12-2015

Genomic Analysis of Sow Reproductive Traits: Identification of Selective Sweeps, Major Genes, and Genotype by Diet Interactions

Melanie D. Trenhaile
University of Nebraska-Lincoln, mtrenhaile@huskers.unl.edu

GENOMIC ANALYSIS OF SOW REPRODUCTIVE TRAITS: IDENTIFICATION OF SELECTIVE SWEEPS, MAJOR GENES, AND GENOTYPE BY DIET INTERACTIONS

By

Melanie Dawn Trenhaile

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Animal Science

Under the Supervision of Professor Daniel Ciobanu

Lincoln, Nebraska

December, 2015

GENOMIC ANALYSIS OF SOW REPRODUCTIVE TRAITS: IDENTIFICATION OF SELECTIVE SWEEPS, MAJOR GENES, AND GENOTYPE BY DIET INTERACTIONS

Melanie D. Trenhaile, M.S.

University of Nebraska, 2015

Advisor: Daniel Ciobanu

Reproductive traits, such as litter size and reproductive longevity, are economically important. However, selection for these traits is difficult due to low heritability, polygenic nature, sex-limited expression, and expression late in life. Marker-assisted selection may provide an alternative to increase genetic progress. Nebraska Index Line (NIL) has been selected for litter size related traits since 1981. It is one of the main contributing lines to the UNL reproductive longevity resource population (n > 1,500), which was genotyped for 60,000 SNPs, phenotyped for age at puberty (AP), lifetime number of parities (LTNP), litter size traits, and other reproductive traits, and fed either a standard or energy-restricted diet during development. These populations are genetic resources to identify polymorphisms associated with reproductive traits as well as polymorphisms that interact with diet to influence reproductive traits that could be used in marker-assisted selection.

Selective sweeps for litter size were detected in NIL using high density genotypes, relative extended haplotype homozygosity, and allelic frequency differences between NIL and its control population, assessed by contingency tests and F_{ST}. Genome-wide association studies (GWAS) for litter size traits identified QTL located next to selective sweep regions identified via multiple methods. These regions harbor potential candidate genes with roles in reproductive processes.

A region on SSC5 that was associated with LTNP and AP was also uncovered by GWAS. The main candidate gene in this region, *AVPR1A*, is associated with social and sexual behavior. Sequencing revealed three non-synonymous SNPs, which were genotyped in 300 individuals with early and late AP. Association analysis indicated linkage between two SNPs and association with LTNP and AP.

Generalized linear mixed models were used to assess the effect of developmental energy restriction and interaction between *AVPR1A* genotype and diet on AP and probability to generate parities 1-3. Energy-restriction delayed AP by 7 d and significantly increased probability of generating parities 2 and 3. Diet and *AVPR1A* genotype interacted to significantly influence parity 3. In addition, GWAS for genotype by diet interaction effects identified eight and four markers that had diet-dependent effects on AP and LTNP, respectively. Single-marker association confirmed the interaction effects between these markers and energy intake prior to breeding.

ACKNOWLEDGEMENTS

First I have to thank my advisor, Dr. Daniel Ciobanu, for sharing his experience, wisdom, stories, and jokes. I have acquired not only knowledge, but also an enhanced ability to think critically and "figure it out" when I don't know the answer right away. Dr. Ciobanu also taught me many life lessons along the way and provided numerous hilarious moments that helped make coming in to work every day really enjoyable. I would also like to thank my committee members, Dr. Merlyn Nielsen, Dr. Rodger Johnson, and Dr. Gary Rohrer for their assistance and support. In addition, I would like to thank Dr. Jessica Petersen for all her help calculating REHH and performing the principle component analysis as well as for being willing to listen and provide life advice. I am very appreciative of Dr. Steve Kachman's statistical assistance with the F_{ST} simulation script and genotype by diet interaction GWAS, among many other things. I am also grateful to Dr. Clay Lents and Dr. Jennifer Thorson for helping with tissue collections, and Dr. Phil Miller and Dr. Tom Burkey for teaching me a lot about nutrition and being really fun to work for.

I would also like to extend my sincere gratitude to Jeff Perkins, Tom McGargill, Kurt Wenninghoff, Cory Willison, Jessica White, and Bob Martin as well as former farm staff at the Swine Research Unit in Mead, NE. They put in so much work every day to make this project happen and are absolutely fantastic people to work with. I would also like to thank Justin Bundy for technical support, and Elizabeth Keuter, Kelsey Timme, Shyamly Premaraj, Hilary Wolf and other undergraduate laboratory staff for all their help. I would also like to express my appreciation to my officemates, Emily Tosky, Katie Lucot, and Taylor Engle, as well as the non-ruminant nutrition students, Shana Barnett, Kelly Moore, Talia Everding, Huyen Tran, Sophie Li, and Dana VanSambeek for their assistance and friendship. I am also thankful for Karina Busato

and all the wonderful people I have met at Christ's Place church who have made my time in Lincoln much more enjoyable.

Finally, I would like to thank my parents, Lance and Michele Trenhaile, brother, Alan Trenhaile, and grandparents, Mary Trenhaile and Sonny and Nyla Blaise for all their support, encouragement, and visits to Lincoln. I am not sure if I could have done this without the encouragement and support of my fiancé, Billy Grannemann, as well as Jillian Yoder and other Missouri friends. I appreciate all the long hours spent and vehicle miles accumulated making many trips to Nebraska and phone conversations when seeing each other wasn't possible. I am sincerely grateful to everyone who helped me achieve my goals, vastly expand my knowledge and abilities, and have a great time in the process here at the University of Nebraska – Lincoln.

Melanie Dawn Trenhaile

TABLE OF CONTENTS

CHAPTER 1: Literature Review

REPRODUCTIVE BIOLOGY

Puberty in gilts is generally defined as the moment of first ovulation (Bidanel, 2011), and occurs between 150 and 220 days of age (Soede *et al.*, 2011), though some early maturing breeds may typically attain puberty as early as 105 days of age (Bidanel, 2011). Many environmental factors, including boar exposure and body condition, influence age at puberty (Soede *et al.*, 2011). Following the onset of puberty, pigs have 18-24 day estrous cycles consisting of two main phases, the follicular phase and the luteal phase (Soede *et al.*, 2011).

The follicular phase lasts four to six days and begins with recruitment of follicles from a pool of antral follicles. Recruitment is stimulated by a change in GnRH secretion from a lesser frequency/greater amplitude to a greater frequency/lesser amplitude pattern. GnRH signals the release of FSH and LH. FSH is important for recruitment, or increase in the number of medium and large sized follicles, and LH is required for selection, or continued growth of some of the recruited follicles to preovulatory size. Selected follicles continue to grow, and upon obtaining sufficient numbers of LH receptors, begin to produce estradiol and inhibin. Inhibin specifically inhibits FSH, causing atresia of the smaller follicles that are still dependent on FSH, while the larger follicles continue to develop. Estrogen levels continue to increase and positively feedback to induce the preovulatory LH surge. Ovulation occurs approximately 30 hours after the peak of the LH surge or 44 hours after the onset of the surge, and the rupture of all follicles takes one to three hours (Soede *et al.*, 2011).

Ovulation rate is affected by many factors. Increased FSH levels may facilitate the recruitment of more follicles, though ample LH is then required for their continued development. IGF1 can aid in the recruitment of follicles by increasing FSH receptors. Nutrition

can play a large role, especially in the event of negative energy balance that often occurs during lactation. Stress can be a factor as prolonged increases in cortisol may delay or even prevent the LH surge by decreasing GnRH secretion (Soede *et al.*, 2011).

Estrus occurs around ovulation and is characterized by receptivity to boars and increased redness and swelling of the vulva as well as mucous production. Estrus is generally detected by the standing response, characterized by immobility, an arched back, and cocked ears in the presence of a boar. The average length of estrus is 40-60 hours but can range from 24 to over 96 hours. Estrus length is affected by intensity of boar contact, stress, parity (shorter in gilts), and wean-to-estrus interval (WEI; shorter at intervals greater than six days). Ovulation occurs about two-thirds of the way through the estrus period (Soede *et al.*, 2011).

Once ovulation occurs, the luteal phase begins. Estrogen and inhibin concentrations decrease as they are no longer being secreted by developing follicles. FSH then increases to recruit a new wave of follicles. Corpora lutea form in place of the ovulated follicles, reaching their full diameter about a week after ovulation. The corpora lutea produce progesterone, which reaches peak concentrations eight to nine days after ovulation. Progesterone inhibits gonadotropin release, keeping follicles small. Factors, such as IGF1, can be important for the formation of luteal tissue and progesterone production. LH becomes important for the support of the corpora lutea after 12 days and is released in a lesser frequency/greater amplitude pattern. If the pig is not pregnant, luteolysis occurs around 15 days post-ovulation and is triggered by prostaglandins secreted by the uterus. Prostaglandins are actually secreted prior to day 15, but the corpora lutea of the pig are not sensitive to them until day 12 or 13. Once the corpora lutea have been lysed, progesterone decreases, and the gonadotropin release pattern transitions to greater frequency/lesser amplitude, beginning the follicular phase (Soede *et al.*, 2011).

Fertilization occurs in the oviduct, and the embryos move to the uterus at the four-cell stage, 60 to 72 hours after the onset of estrus. They reach the blastocyst stage by day five and quickly elongate by day 16 in order to have maximum surface area for contact between the trophectoderm and uterine luminal epithelium for the uptake of nutrients. The corpora lutea must be maintained throughout the pregnancy in order to secrete progesterone (Bazer and Johnson, 2014). Prostaglandins that lyse the corpora lutea in non-pregnant pigs are still secreted, but estrogen from the blastocysts signal pregnancy recognition and cause the prostaglandins to be secreted into the uterine lumen rather than into circulation so they do not reach the corpora lutea to cause lysis (Soede *et al.*, 2011). Secretions from the uterine luminal and glandular epithelium are also required to support attachment, development, and growth of the embryos, and cellular remodeling at the uterine luminal epithelium trophectoderm interface is necessary during implantation. Implantation in the pig is non-invasive and eventually results in the development of an epitheliochorial placenta (Bazer and Johnson, 2014). Pregnancy lasts for 114 to 116 days (Soede *et al.*, 2011), and the farrowing process takes three to five hours on average (Bidanel, 2011). Parturition is initiated by cortisol production from the mature fetal hypothalamic-pituitary-adrenal axis. Uterine contractions and cervical pressure from the fetus trigger oxytocin release from the posterior pituitary, which then leads to large pulses of prostaglandins to regress the corpora lutea and stop progesterone production. Prostaglandins also help mediate the processes of placental membrane rupture, cervix softening and dilation, myometrial contractions, separation of placental membranes from the uterus, expulsion of piglets, and uterine involution (Bazer and Johnson, 2014).

Fertilization rates typically approach 100%, and prenatal mortality generally ranges from 30-40%. The average litter size is around 14.5 for maternal breeds, such as Large White and Landrace, and 9.9 for paternal breeds, such as Duroc. Up to ten percent of these piglets may be

stillborn in some populations. During lactation, sows have very little follicular development and do not ovulate or express estrus. Upon weaning, sows will generally ovulate within four to ten days (Bidanel, 2011). The weaning to estrus interval can be longer in sows with negative energy balance. Negative energy balance may also reduce LH pulses, which leads to fewer recruited and selected follicles and a lower ovulation rate (Soede *et al.*, 2011).

GENETIC IMROVEMENT OF REPRODUCTIVE TRAITS

Genetic Variation and Heritability

Female reproductive traits are mostly lowly heritable, though a few fall in the moderate range. Fertility and prolificacy traits tend to be very lowly heritable due to complex interactions between sow, boar, and embryo genotypes (Bidanel, 2011) in addition to substantial environmental influence. Traits such as age at puberty (AP), ovulation rate (OR), and WEI may have moderate heritabilities since the genetic variation influencing these traits is a result of the sow's own genotype (Bidanel, 2011). Heritability estimates for reproductive traits differ widely between studies and breeds. Heritability of AP has been estimated at 0.19 (Schneider *et al.*, 2011), 0.29 (Knauer *et al.*, 2010a; Knauer *et al.*, 2011), 0.38 (Tart *et al.*, 2013), and 0.57 (Hsu, 2011). Heritability for other estrus traits, including strength and length of standing response and vulva redness and swelling, range from 0.2 to 0.57 (Knauer *et al.,* 2010a; Knauer *et al.*, 2011). Total number born (TNB) heritability estimates vary from 0.02 to 0.16 (Knauer *et al.*, 2011; Schneider *et al.*, 2012a; Tomiyama *et al.*, 2011; Tart *et al.*, 2013), while number born alive (NBA) estimates range from 0.02 to 0.2 (Abell *et al.*, 2013; Chen *et al.*, 2003; Schneider *et al.*, 2012a; Schneider *et al.*, 2011; Tart *et al.*, 2013; Hsu, 2011). Lifetime NBA heritability was estimated at 0.12 by Abell *et al.* (2013). Number weaned (NW) heritability has been estimated to be between 0.05 and 0.23 (Chen *et al.*, 2003; Schneider *et al.*, 2011; Hsu, 2011). Wean-to-estrus interval heritability was estimated at 0.02 (Schneider *et al.*, 2011), and embryonic survival heritability was estimated at 0.17 (Hsu, 2011). Heritability of OR was estimated at 0.27 (Hsu, 2011) and 0.45 (Schneider *et al.*, 2011), while litters/sow/year heritability was estimated at 0.11 (Abell *et al.*, 2013). Sow reproductive longevity heritability estimates also exhibit a wide range of values

across studies, breeds, and specific definitions used. These estimates vary from 0.02 to 0.25 (Serenius and Stalder, 2006; Tart *et al.*, 2013; Mészáros *et al.*, 2010; Knauer *et al.*, 2011).

Additive genetic variance for NBA (Schneider *et al.*, 2011; Schneider *et al.*, 2012a; Abell *et al.*, 2013) and NW (Schneider *et al.*, 2011) was present in various populations of pigs, though values varied between breeds (Chen *et al.*, 2003). Additive genetic variance was also found for TNB (Schneider *et al.,* 2012a), WEI (Schneider *et al.*, 2011), OR (Schneider *et al.*, 2011), AP (Schneider *et al.,* 2011), and reproductive longevity (Mészáros *et al.,* 2010).

In addition to within breed variation, much variation exists between breeds for reproductive traits. Much emphasis has been placed on reproductive traits in certain breeds, such as Large White and Landrace, which have become known for superior maternal performance. On the other hand, the selection criteria for other breeds, such as Duroc and Hampshire, have focused on lean growth and carcass traits. For example, TNB averages 14.2, 14.6, and 9.9 for Large White, Landrace, and Duroc, respectively (Bidanel, 2011). Despite selection for similar traits, there are differences in reproductive performance between common commercial maternal breeds (Bidanel, 2011) as well as genetic lines originating from commercial maternal breeds (Serenius *et al.*, 2006; Knauer *et al.*, 2010b). For example, Mészáros *et al.* (2010) compared reproductive longevity between Large White and Landrace sows, and found that Landrace sows complete 0.56 more parities on average, while Large White sows have approximately 0.5 more piglets per litter. There are also many local breeds that differ widely in reproductive traits. However, none compare to common commercial maternal breeds, except the Chinese Meishan breed which has been known for superior reproductive performance, though recent selection in commercial maternal breeds have nearly increased their performance to the level of the Meishan breed (Bidanel, 2011). In a study by Canario *et al.* (2006), Meishan TNB was only one piglet higher than Large White TNB. Meishans, however, still

have higher conception rates and prenatal survival at a given ovulation rate than Large White. They also reach puberty up to 100 days earlier than Western breeds and have shorter WEI (Bidanel, 2011).

Though estimates are quite variable between populations and calculation methods used, ample genetic variation and heritability exists for effective selection for sow reproductive longevity (Serenius and Stalder, 2006; Serenius *et al.*, 2006; Knauer *et al.*, 2010b) and litter traits (Tomiyama *et al.*, 2011) as well as age at puberty and estrus traits (Knauer *et al.*, 2010a).

Phenotypic Selection

Despite low heritability, selection experiments for reproductive traits have shown that progress can be made through traditional selection. Zimmerman and Cunningham (1975) performed a five generation selection experiment for increased OR, as OR is one of the main components of litter size. By performing laparotomies nine to eleven days following second estrus and selecting gilts with higher ovulation rates, they were able to increase the average OR from 14.38 to 16.19 in five years. The control line, which was derived from the same population at the same time, started the experiment with an average ovulation rate of 14.63 and ended with an average ovulation rate of 13.67. Female selection was performed at random in the control line, and male selection and matings were random in both lines. The average unweighted selection differential per generation was 1.18 corpora lutea. Regression of response on the cumulative selection differential generated a realized heritability estimate of 0.48.

Bennett and Leymaster (1990b) simulated selection for litter size via direct selection for litter size and selection for OR, uterine capacity (UC), and ES, which are components of litter size. In addition, two indexes including OR and UC, an index including OR and ES, and an index including OR and litter size were simulated. The simulation results indicated progress could be

made with any of these selection methods. However, index selection was more successful than direct selection, with the OR/UC indexes producing a 37% greater response and the OR/ES and OR/litter size indexes producing a 21% greater response than direct selection for litter size. Direct selection for litter size was more effective than direct selection for any litter size component alone. Direct selection for ES produced the lowest response, and the index of UC and OR that was weighted to increase UC at 95% of the increase in OR produced the highest response. While these are simulations rather than experiments involving real animals, they still provide a valuable comparison of potential selection methods and strengthen the argument that selection for litter size is possible.

Lamberson *et al.* (1991) selected for increased OR for nine generations in a 14 breed composite population. After two generations of random selection, the selection line was randomly split into three lines. The first line was selected for increased litter size, the second line was selected for decreased AP, and the third line continued random selection for eight generations. After the first nine generations of selection, OR increased by 3.7 eggs. Selection for OR increased litter size by 0.089 pigs per generation. Only 20% of the additional ova that were ovulated as a result of selection for OR actually produced an additional live born piglet; thus, selection for increased OR is not the most effective way to increase litter size. At the end of the experiment, the randomly selected line maintained 75% of the total response in OR achieved after nine generations of selection, suggesting that part of the increase in OR was due to selection for gene combinations (epistasis). Continued selection is required to prevent recombination from breaking up the gene combinations that were under selection. The total response to selection for litter size at the end of the experiment was 1.8 pigs per litter when the regression method was used and 1.4 pigs per litter when the animal model was used. Total response for AP was -15.7 and -17.1 days when regression and the animal model were used,

respectively. Selecting for decreased AP had no effect on litter size. This experiment demonstrates that OR, litter size, and AP can all be improved via selection.

Johnson *et al.* (1999) practiced index selection for increased OR and ES for 11 generations, then selected for increased TNB for three generations. After the first 11 generations, the number of ova and fetuses at day 50 increased by 7.4 and 3.8, respectively. Total number born increased by 2.3 pigs and NBA increased by 1.1 pigs. Responses to selection at the end of the experiment (generation 14) were 3.0 and 1.4 pigs for TNB and NBA, respectively. While these increases due to selection were statistically significant, NW declined (though not significantly). Because of this and the positive genetic correlation between OR and number of fetuses and number of stillborn pigs (SB) and mummies (MUM) per litter, it would be more effective to emphasize NBA rather than TNB and include selection for increased birth weight to produce more live, viable piglets.

Bolet *et al.* (2001) performed selection for increased litter size for 17 generations. Selection was performed within sire family on number of piglets born during a sow's first two parities during the first ten generations. In generations 11-17, selection was performed within-sire for males and across-sire for females based on number of piglets born in parity 1 only. These changes in selection strategies increased overall selection intensity by 79 percent due to increased population size as more sows that had produced one litter were available for selection than sows that had produced two litters. Beginning in generation 11, the line was also opened to daughters of hyperprolific boars and sows, with an immigration rate of ⅛ each generation. Total number born in parities one and two and ova shed and number of live embryos at 30 days during the third pregnancy all exhibited significantly positive BLUP responses per generation for all generations. However, the response to selection was between three and four fold higher after generation 11. Overall, the gain in litter size (averaged over parities one and two) was 1.4

piglets per litter. The authors estimated that a gain of 0.8 piglets per litter was due to immigration and a gain of 0.6 piglets per litter was due to within line selection.

Noguera *et al.* (2002) conducted a selection experiment for NBA using a BLUP repeatability animal model. Genetic response was analyzed using Bayesian methods and a multivariate model carried out by Gibbs sampler. This experiment used both family information and high selection intensity as the population selected from was quite large. After one generation of selection, posterior means of standardized selection differentials for selection line females in parities one through six ranged from 0.70 in parity six to 0.94 in parity four. Posterior means of standardized selection differentials for selection line males ranged from 0.22 in parity six to 0.34 in parity four. The highest posterior density regions of 95% did not include zero in all but parity six selection males, indicating that selection was effective in both males and females; however, it was more effective in females since sows were selected on their own information as well as relative information, whereas males could be selected based on relative information only. The posterior means of standardized selection differentials for the control line were slightly positive; however, zero was included in the 95% highest posterior density region for all parities, indicating that selection really was random. The posterior means of direct genetic response ranged from 0.32 in parity one to 0.64 in parity six. The results from this experiment indicate that selection for litter size is most effective when a family selection index is used along with intense selection in a large population. These results also suggest that litter size in each parity may have different genetic influences.

Holl and Robison (2003) conducted a nine generation experiment in which they selected based on estimated breeding values (EBVs) for NBA in order to increase litter size. Cross-fostering was performed within 24 hours, so that each gilt was reared in a litter of ten piglets or less. Genetic trends were 0.053 pigs/year and phenotypic trends were 0.145 pigs/year in the

selection line. The selection line had higher cumulative selection differentials, EBVs, and litter sizes than the control line by 9.05, 0.63, and 0.86 pigs, respectively. This study demonstrates that litter size can be improved using direct selection for NBA using breeding value estimates when gilts are nursed in litters of ten piglets or less.

Nielsen *et al.* (2013) reported the results of selection for litter size five days after farrowing (LS5) in Danish Landrace and Yorkshire pigs. Genetic correlations for Landrace and Yorkshire between TNB and mortality, TNB and LS5, and mortality and LS5 were 0.28 and 0.22, 0.74 and 0.68, and -0.43 and -0.57, respectively. Phenotypic correlations for Landrace and Yorkshire between TNB and mortality, TNB and LS5, and mortality and LS5 were 0.14 and 0.09, 0.77 and 0.71, and -0.47 and -0.59, respectively. The unfavorable correlations between TNB and mortality demonstrate that selection for TNB generally results in increased mortality. However, LS5 has a favorable correlation with both traits, and selection for LS5 can both increase TNB and reduce mortality. Selection for LS5 from 2003 to 2009 resulted in genetic improvements of 1.7 piglets for LS5, 1.3 piglets for TNB, and -4.7% mortality in Landrace and 2.2 piglets for LS5, 1.9 piglets for TNB, and -7.9% mortality in Yorkshire; phenotypic improvements were 1.4 and 2.1 piglets for LS5, 0.3 and 1.3 piglets for TNB, and -7.9 and -7.6% for mortality in Landrace and Yorkshire, respectively. While mortality remained slightly greater in the litters with larger TNB as compared to litters with medium and small TNB, the trends for genetic and phenotypic change were nearly identical for litters of all sizes; therefore, selection for LS5 reduced piglet mortality in large as well as medium and small litters.

Marker-assisted selection, genome-wide association studies, and QTL mapping

Phenotypic selection for reproductive traits can be difficult due to low heritability, intensive measurement, sex-limited nature, and expression late in life. For example, sow

reproductive longevity is a trait with high importance to the swine industry, yet it is rarely selected for as it is very lowly heritable and expressed throughout the entire life of a sow. While AP is an early indicator of this trait, selection for AP is also not practiced in industry herds due to the tedious, daily, labor-intensive data collection required (Tart *et al.*, 2013). Advances in the fields of molecular biology and genomics, particularly the development of marker panels spanning the entire genome, make quantitative trait locus (QTL) identification and genomic selection based on an overall sum of marker effects or marker-assisted selection (MAS) based on known marker-QTL associations (Uimai *et al.*, 2011) a feasible solution to these problems. (Meuwissen *et al.*, 2001; Bidanel, 2011). While phenotypic measurements for most reproductive traits are not available until after replacement gilts are selected, genomic marker results can be available at a young age. Information from markers can be combined with relative information or used alone in order to select both females and males early in life (Schneider *et al.*, 2012a). In addition to decreasing generation interval and allowing the selection of both sexes, the inclusion of genetic markers in the information used to make selection decisions may also increase the accuracy of selection and, therefore, the selection response of the trait (Meuwissen *et al.*, 2001; Spötter and Distl, 2006).

Many studies have identified QTL and/or candidate genes for reproductive traits. Currently, 12,618 QTLs from 461 publications are referenced in the PigQTLdb (http://www.animalgenome.org/cgi-bin/QTLdb/SS/index). Spötter and Distl (2006) review two major methods for identification of markers to be targeted for selection. The first method is the candidate gene approach. Genes may be identified as candidates for influencing a trait due to their physiological roles relating to the trait of interest (physiological candidate genes), location near an identified QTL as well as having an orthologous gene that is known to influence the trait of interest (positional candidate genes), and differential expression in the tissue under

investigation. Once a candidate gene is identified, polymorphisms are found within the gene and association studies are performed to provide evidence that the gene is indeed associated with the trait of interest (or a marker for a closely-linked associated gene or regulatory region). This method has been successfully employed to identify genes associated with litter size by many researchers, including Rothschild *et al.* (1996), who identified a *PvuII*-restriction fragment length polymorphism (RFLP) in estrogen receptor 1 (*ESR1*), Vincent *et al.* (1998), who identified a polymorphism in prolactin receptor (*PRLR*), Vallet *et al.* (2005), who identified a polymorphism in erythropoietin receptor (*EPOR*), Short *et al.* (1997b) and Hamann *et al.* (2000), who found significant effects in some of the 13 alleles present at a microsatellite marker linked with osteopontin (*OPN*), Li *et al.* (1998), who identified a marker within follicle-stimulating hormone β (*FSHB*), and Spötter *et al.* (2005), who found a polymorphism in leukemia inhibitory factor (*LIF*). Later studies failed to confirm some of these associations in other populations because the presence or absence of an association varies between breeds, sample sizes, and populations due to differences in linkage, recombination, and epistatic effects (Spötter and Distl, 2006).

The second major method for identifying markers to be targeted for selection is QTL mapping (Spötter and Distl, 2006). This method seeks to identify genomic locations associated with traits of interest. Specific goals may include understanding the magnitude of loci effects and/or the form of gene action (additive, dominance, and epistasis) as well as precisely locating markers linked to causative mutations that affect the trait of interest or even the causative mutations themselves. Quantitative trait locus mapping can be performed on a reduced set of polymorphisms or on a genome-wide basis using thousands or even millions of polymorphisms located across the genome (Jamann *et al.*, 2015). Spötter and Distl (2006) list 18 QTL for OR, SB, TNB, UC, and prenatal survival, located on 11 of the chromosomes in the pig genome discovered via QTL mapping. Most of these QTL have suggestive linkage; however, three reach significant

linkage. PigQTLdb currently lists 138, 130, and 130 QTL for TNB, NBA, and OR, respectively. Several QTL have also been identified for total litter weight at birth (n = 39), SB (n = 77), and MUM (n = 95), and 210 QTL have been identified for AP.

Perhaps the most useful approach is to merge the two methods via first identifying and fine mapping a QTL, then locating physiological candidate genes in the chromosomal region of the QTL with the ultimate goal of identifying the gene and specific polymorphism responsible for the QTL. Despite the large number of QTL identified, there has been a lack of success in identifying functional polymorphisms that are the source of the QTL effect. There are several explanations of why there may not be any physiological candidate genes located near QTL or QTL located near physiological candidate genes. First, if the effect of the gene is small enough, it may not be detected by genome-wide QTL searches. Second, the alleles in the candidate gene may not be segregating in certain populations. In addition, the marker and causative mutation may not be in complete linkage disequilibrium (LD), and candidate gene association effects may have been due to chance or QTLs not found due to low statistical power (Spötter and Distl, 2006). Pomp *et al.* (2001) hypothesized that QTLs typically represent regulatory elements or factors that initiate a cascade of events that influence the expression of physiological genes rather than the physiological genes themselves.

Identifying differentially expressed genes can be a useful tool for identifying genes associated with reproductive traits. Techniques, such as microarray technology and RNAseq, make the simultaneous analysis of the expression of thousands of genes or the entire transcriptome possible. Several genes that have been investigated as candidate genes were found to be differentially expressed in the fetus or reproductive tissues, strengthening their argument as candidate genes (Spötter and Distl, 2006).

Rempel *et al.* (2010) sequenced five candidate genes for reproductive traits and selected 53 SNPs to be genotyped along with other SNPs previously identified in the literature in a population of swine comprised of Yorkshire, Landrace, and Duroc breeds. In total, 76 SNPs were genotyped, and association analyses were performed for AP, OR, WEI, TNB, NBA, SB, and MUM. In total, 11 SNPs were found to be associated with AP in six genes. Six SNPs in five genes were associated with OR. Wean-to-estrus interval had four and five associated genes and SNPs, respectively. Litter size traits had fewer associated SNPs; two genes and three SNPs were associated with TNB, and both genes were also associated with NBA. Number of stillborn piglets had three and four associated genes and SNPs, respectively, and one SNP was weakly associated with MUM. Many of the genes associated with reproductive traits in this study have functions relating to metabolism, suggesting an important link between reproductive efficiency and energy utilization. However, the candidate SNPs tested are likely just markers linked to causative genetic variation and are not actually influencing these traits themselves.

Selection for increased litter size has been successful in Danish pig breeding. Landrace sows with low and high EBVs for litter size from the Danish population were genotyped for 17 microsatellite markers in regions corresponding to previously identified QTLs for litter size on chromosomes 11, 13, and 15. Most of the sows (~90%) could be assigned to the correct EBV group based on genotype information. The average identity by state relationship was greater in the region on chromosome 13 than other regions, indicating that selection had targeted this region. Even so, sufficient genetic variation still exists to target this region in marker-assisted selection. Several markers in all three investigated regions carry considerable variation in allele effects, so marker-assisted selection will be most effective if several markers are used (Bjerre *et al.*, 2010).

Genetic improvement has also been made for reproductive traits in the Finnish Landrace pig population. Uimari *et al.* (2011) identified five single nucleotide polymorphisms (SNPs) significantly associated with TNB parity 1 and parities 2 and later, two SNPs suggestively associated with SB in parities 2 and later, two SNPs suggestively associated with piglet morality between birth and weaning in parity 1, three SNPs significantly and three SNPs suggestively associated with piglet morality in parities 2 and later, one SNP suggestively associated with first farrowing interval, and one SNP suggestively associated with second farrowing interval in this population. Chromosome 9 is home to significant SNPs for TNB at 79 and 95 Mb. Three SNPs near 79 Mb were in complete LD, and these three SNPs were in moderate disequilibrium with the significant SNP at 95 Mb. Several candidate genes with physiological functions relating to reproduction are located in these regions. Several SNPs located around 66 Mb on chromosome 9 are associated (some significantly and others suggestively) with piglet mortality. These SNPs are in strong LD with each other, but not with any of the SNPs associated with TNB.

Onteru *et al.* (2011) used a Bayes C approach to perform GWAS in commercial maternal sows for reproductive traits, including TNB, NBA, SB, MUM, and gestation length (GL), in each of the first three parities with Porcine SNP60 BeadArray (Illumina) genotypes. Total number born, NBA, and MUM in the first three parities were lowly heritable, while SB and GL in the first three parities were moderately heritable. Consequently, only a small proportion of the phenotypic variance was explained by the genomic markers (0.001-0.40). Phenotypic correlations for these traits across the first three parities were low: 0.16-0.2 for TNB, 0.15-0.18 for NBA, 0.09-0.21 for SB, 0.009-0.06 for MUM, and 0.33-0.38 for GL. Candidate QTL regions were sliding five SNP windows that had a proportion of variance explained greater than $8.6*10^{-5}$ (an expected proportion of variance accounted for by a window) and a bootstrap analysis P-value < 0.01 (0.05 for MUM). The regions meeting these criteria were mostly different between parities for each

trait, further suggesting that each parity is influenced by different genes and should be considered separate traits. For all traits except MUM, an average of 68.3 percent of all genes present in the QTL regions were reproductive genes involved in pituitary, ovarian, uterine, placental, and embryological functions, while an average of 48.9 percent of all genes present in the QTL regions were involved in placental function. Pathway analysis discovered that nucleotide metabolism pathways were enriched in the genes located in SB QTL regions for all three parities, suggesting that adding nucleotides to pig diets may reduce stillborn piglets. While many promising candidate QTL regions, genes, and pathways were identified, validation studies are still needed for confirmation due to large environmental influence and the lowly heritable, polygenic nature of reproductive traits.

Schneider *et al.* (2012b) performed a GWAS for several parity one litter size traits using genotypes from Landrace-Duroc-Yorkshire dams and the Porcine SNP60 BeadArray. The SNPs were grouped in sets of five consecutive SNPs by chromosome position for analysis using a Bayes C model. Bootstrap analysis was used for hypothesis testing, and after corrections for multiple testing, 124 statistically significant ($P < 0.01$) QTL were identified. Eleven of these QTL were for TNB on SSC 1, 4, 13, 14, 15, and 17, 14 were for NBA on SSC 1, 4, 6, 10, 13, 15, and 17, one was for number born dead (NBD) on SSC11, 33 were for litter birth weight (LBW) on SSC 1, 2, 3, 4, 5, 6, 7, 9, 10, 14, 15, and 17, and 65 were for average birth weight (ABW) on SSC 1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, and 18. No QTL were identified for SB and MUM. Five, six, one, and more than eight candidate genes were identified in QTL regions for both TNB and NBA, NBA only, NBD, and ABW, respectively. These QTL may be of use in marker-assisted selection, marker-assisted management, or genomic selection, and may be enhanced by fine mapping and sequencing to identify potential causative polymorphisms.

Tart *et al.* (2013) performed GWAS with a Bayes B approach to identify 1-Mb windows that influenced AP and sow reproductive longevity as AP is moderately heritable and an early indicator of reproductive longevity. Gilt that express first estrus early in life tend to have improved reproductive longevity. Several of the QTL regions identified overlap with QTL regions identified by others in previous studies. Genes located in the top one percent of 1-Mb windows were enriched for regulation of small GTPase-mediated signal transduction, positive regulation of vasoconstriction, and RAS protein signal transduction ($P < 0.05$). Presence or absence of expression of six positional candidate genes, *AVPR1A*, involved in reproductive behavior, *BAIAP2,* which plays a role in neuronal growth, *CRTC1,* whose functions include regulation of transcription, *OR2G3,* which is involved in detection of chemical stimulus, *PAPPA,* which influences pregnancy, and *PRKAA2*, which plays a role in energy metabolism, were analyzed in the hypothalamus, pituitary and ovarian cortex of prepubertal gilts. All six genes were expressed in the ovarian cortex, and all genes but *PAPPA* were detected in the pituitary. Only *AVPR1A* and *OR2G3* were expressed in the hypothalamus. Non-synonymous SNPs were located in *AVPR1A* and *BAIAP2*, and synonymous SNPs were located in *AVPR1A, OR2G3,* and *PAPPA.* Single-marker association was used to evaluate the SNP with the largest effect on AP in each of the top one percent of QTL windows identified by GWAS. Ninety-two percent of these 26 SNPs were associated with significant differences in AP with an effect size ranging from 3.4 to 11.1 days of age between opposite homozygotes ($P < 0.05$). Two of these SNPs, including a non-synonymous SNP located in *AVPR1A*, were also associated with LTNP ($P < 0.05$). Linkage disequilibrium in this region is high, and SNPs in this region may be useful markers to employ in marker-assisted selection for AP and reproductive longevity.

Lillehammer *et al.* (2013) performed simulations to determine the optimal genomic selection scheme under a scenario similar to that of the Norwegian Landrace. A production trait

and a maternal trait, representing indices of production and maternal traits, were simulated. Heritability was assumed to be 0.1 for maternal traits and 0.3 for production traits, and they had a negative genetic correlation of -0.3. Four times per year, the 300 dams and 25 boars with the highest breeding values were mated to produce selected litters. Two females and one male were randomly selected from each litter. Zero, 1200, or 2400 females and all 1200 selected males were genotyped each year. Alternatively, within litter selection was also used. This was accomplished by genotyping two males from each litter (2400 total) and zero, 1200, and 2400 females. In each scenario, genomic selection resulted in greater genetic gain. Genetic gain increased and inbreeding decreased as more females were genotyped. Within litter selection increased genetic gain, but also increased inbreeding. The percent of genetic gain due to the maternal trait increased as the number of females genotyped increased, but using within-litter selection didn't have an appreciable effect. The schemes ranked the same when the correlation between traits was zero and heritability of the production trait was increased to 0.4 or decreased to 0.2. The genetic gain was the same or greater with uncorrelated traits rather than negatively correlated traits, and the difference between schemes was smaller. Increasing production heritability increased genetic gain and decreased the percent of genetic gain due to the maternal trait. Cost-benefit analysis showed that all tested schemes would be beneficial over conventional selection. While regular genomic selection (only genotyping 1200 males) where 2400 females are genotyped gives similar genetic gain to within-litter selection (genotyping 2400 males) where 1200 females are genotyped, genotyping more females rather than more males is advisable as it results in a lower rate of inbreeding as well as a balanced relative genetic gain of the two traits.

Tao *et al.* (2013) identified *TCF12, CTNNAL1,* and *WNT10B* as candidate genes for litter size due to their proximity to known QTL for litter size traits, functions related to reproduction,

and differential expression. These genes are involved in the WNT signaling pathway, which is essential for reproductive system development and affects processes such as follicular development, ovulation and luteinization of ovarian follicular cells, and establishment of pregnancy. Four SNPs (two in *TCF12*, one in *CTNNAL1*, and one in *WNT10B*) were identified. The two SNPs in *TCF12* were in complete LD in Large White and moderate LD in DIV, a population consisting of Landrace, Large White, Tongcheng, and Taihu breeds. Association analyses were performed for these SNPs and NBA and TNB in parity one and later parities in Large White and DIV pigs. One SNP in *TCF12* and the SNP in *CTNNAL1* were significantly associated with litter size traits in these two lines of pigs, and have the potential to be causative mutations of nearby QTL for litter size.

Nonneman *et al.* (2013) performed a GWAS for delayed puberty using genotypes from the Porcine SNP60 BeadArray in 91 gilts that did not display estrus by 240 days and 127 pubertal littermates. The top 174 SNPs with the highest associations were genotyped in an additional 86 non-pubertal and 103 pubertal gilts. Twelve of these SNPs were found to be significant, and candidate genes were identified near some of them. The two most significant SNPs were located on SSC4 surrounding *NHLH2*, a gene that has been associated with delayed puberty in mice, possibly due to downregulation of GnRH receptors.

HETEROSIS AND CROSSBREEDING

The production system used most commonly by the swine industry in the United States and many other countries can be described with a three-tiered pyramid. The top of the pyramid is the nucleus. This level contains the smallest proportion of animals and is where the genetic improvement occurs (Abell *et al.*, 2010). The nucleus farms are often comprised of purebred animals, but many swine genetic companies develop their own synthetic lines by crossing two or more breeds. Breeds such as Duroc and Piétrain are typically used as paternal lines and are selected for finishing efficiency and meat quality traits. Breeds such as Large White, Landrace, and Yorkshire are usually used as dam lines and are selected for sow productivity (Dekkers *et al.*, 2011). The second level of the pyramid is the multiplier, where the genetic progress from the nucleus is multiplied to produce a larger number of animals that can go into the bottom tier of the genetic pyramid, the commercial level. This level represents the majority of the animals in the system, and produces pigs with the purpose of going to market to produce pork (Abell *et al.*, 2010).

Johnson and Omtvedt (1975) compared the performance of purebred and crossbred gilts for reproductive traits. There was very little difference between the two groups in reproductive failure rate. Crossbred gilts of all breed groups had fewer corpora lutea 30 days post-breeding than purebred gilts (13.48 vs. 13.03); however, the embryo survival rate of crossbred gilts was 9.52±3.23% higher than purebred gilts. Crossbred gilts also had larger and heavier litters at all ages and had a larger percentage of their litter survive from birth to weaning than purebred gilts. Similarly, Wilson and Johnson (1981) found that three-breed crosses of Duroc, Hampshire, and Yorkshire produced larger and heavier litters at all ages than backcrosses

consisting of the same breeds. The three-breed crosses also gained faster and utilized feed more efficiently than backcrosses.

Kuhlers *et al.* (1994) studied differences between Duroc, Yorkshire, and Landrace three-breed rotational and terminal crossbreeding systems. Rotational crossbreeding systems can be advantageous due to production of replacement gilts in the system. However, it has several disadvantages including reduced ability to capitalize on breed complementarity, increased intergenerational variability, and increased management difficulty. In addition, not all heterosis is realized in rotational crossbreeding systems, whereas terminal crossbreeding systems capture 100% of the potential heterosis. Litter sizes were numerically greater in the terminal system than the rotational system at all ages, but the difference was not significant until market age. Breed composition influenced the variation of all litter size traits, except number marketed. The terminal crossbreeding system also produced significantly heavier litters at birth, 56 days, and market age; however, litter weight was statistically the same at 21 days. Breed composition affected litter weights at birth and 21 days. There were no differences between the two crossbreeding systems for sow weight and backfat thickness (BF) at weaning, lactation feed intake, and farrowing rate. However, terminal crossbreeding system sows had shorter WEI. Breed composition of the sow did have an effect on sow weight and BF at weaning, WEI, and AP.

Cassady *et al*. (2002) performed two experiments to study the effects of heterosis. Yorkshire, Landrace, Large White, and Chester White breeds were used in experiment 1, and Duroc, Hampshire, Piétrain, and Spot breeds were used in experiment 2. Data were recorded on purebred and two-breed cross pigs as well as generations F_1 – F_6. In general, gilts in experiment 1 were more productive carrying crossbred rather than purebred litters, though the opposite was true for gilts in experiment 2. Crossbred gilts reached puberty sooner, weighed more at farrowing, and produced larger and heavier litters than purebred gilts. Direct heterosis

significantly decreased AP and increased sow weight at 110 days of gestation and litter weight at day 14 and at weaning in both experiments. In experiment 1, direct heterosis also decreased gestation length. In experiment 2, direct heterosis increased number of nipples, weight at puberty, lactation weight loss, litter size, and LBW. Maternal heterosis decreased sow weight at 110 days of gestation in experiment 1 and increased AP in experiment 2. In experiment 1, litter heterosis significantly increased number of pigs at 14 days, weaning weight, and litter weights at birth, 14 days, and weaning. In experiment 2, litter heterosis decreased litter size (Cassady *et al.*, 2002). Utilization of crossbreeding systems can maximize productivity in swine herds by capitalizing on heterosis and breed differences (Kuhlers *et al.*, 1994).

RELATIONSHIP BETWEEN REPRODUCTIVE TRAITS AND OTHER ECONOMICALLY IMPORTANT TRAITS

Relationship between Reproductive Traits and Developmental Traits

Noguera *et al.* (2002) used Bayesian analysis to look at direct genetic response to selection for litter size and correlated response in weight and BF at 175 days of age via a multivariate model carried out with a Gibbs sampler. The posterior means of correlated genetic responses were small at -0.66 kg for weight and 0.2 mm for BF, and the 95% highest posterior density regions contained zero correlated responses. Holl and Robison (2003), however, found genetic correlations of 0.10 and 0.54 between NBA at parity 1 and days to 104 kg and adjusted BF to 104 kg, respectively, in a Large White – Landrace composite population that was selected for NBA. In the control line, there was no genetic correlation between NBA and days to 104 kg, and the genetic correlation between NBA and adjusted BF was smaller at 0.36.

Holm *et al.* (2004) estimated genetic correlations (r_g) in Norwegian Landrace females between adjusted age at 100 kg and age at first service (AFS), NBA parity 1 (P1) and parity 2 (P2), and wean-to-service interval (WSI) after P1 and P2 at 0.68, 0.60, 0.42, 0.16, and 0.20, respectively, suggesting an unfavorable correlation between growth rate and reproductive traits. Likewise, genetic correlations between feed consumption and AFS and NBA-P1 and -P2 were also unfavorable, with r_g of 0.20, 0.23, and 0.20, respectively. The authors hypothesized that selection for growth results in the necessity for gilts to put all their energy into growth rather than reproduction, whereas selection for reproduction would cause the energy to be partitioned predominantly to reproduction rather than growth. In general, adjusted BF was not correlated with any reproduction traits. However, Tummaruk *et al.* (2001) generated conflicting

results when they studied the effects of several developmental traits on litter size and other reproductive traits in parities 1-5 of Swedish Landrace and Yorkshire nucleus sows. In all parities (1-5), sows that had a higher growth rate of up to 100 kg body weight had larger litter sizes and shorter WSI than sows with a lower growth rate ($P < 0.05$). High BF at 100 kg resulted in shorter WSI at P1 ($P < 0.001$) and increased litter size in P2 ($P < 0.01$). Furthermore, Serenius *et al.* (2004) found favorable genetic correlations between average daily gain (ADG) and SB in Finnish Landrace (-0.38) and Large White (-0.25) pigs. Favorable genetic correlations were also present between feed to gain ratio and SB in both breeds (0.27 in Landrace and 0.12 in Large White).

In Finnish Landrace and Large White crossbred pigs, Serenius and Stalder (2007) found positive associations between length of productive life and age at 100 kg live weight ($P < 0.05$). However, they were unable to find an association between BF thickness adjusted to 100 kg live weight and length of productive life. Hoge and Bates (2011), however, found a negative association between growth and sow longevity and a positive association between BF thickness and sow longevity in commercial Yorkshire sows, as fatter, slower growing sows had a decreased risk of being culled. Sobczyńska *et al.* (2013) were unable to find genetic correlations between ADG or BF thickness and sow longevity in Polish Landrace and Large White sows. However, the phenotypic selection index, which combines growth and BF as well as *longissimus* muscle depth, was significantly and unfavorably correlated with several longevity traits, including length of life (defined as days from birth to culling), number of litters, and lifetime pig production in both breeds. Significant unfavorable correlations were also present between phenotypic selection index and lifetime litter efficiency (total number litters / days between first and last farrowing) and lifetime pig efficiency (lifetime pig production / days between first and last farrowing) in Landrace sows. Serenius *et al.* (2006) found greater ADG and greater BF decreased risk of being culled in two and five, respectively, of the six genetic lines they

evaluated. Knauer *et al.* (2010b) found a negative correlation between ADG and stayability to parity 4 (STAY4) and a positive correlation between BF and STAY4 in three of six commercial maternal genetic lines evaluated. Due to these mixed results, the effect of ADG and BF on sow longevity may be population dependent.

Relationship between Reproductive Traits and Conformation Traits

Serenius *et al.* (2004) investigated genetic correlations between prolificacy traits and conformation traits in Finnish Landrace and Large White pigs, but no clear associations were identified. Serenius and Stalder (2007), however, did find that leg soundness did significantly influence length of productive life in Finnish Large White and Landrace crossbred pigs ($P < 0.001$), likely due to an increased lifespan in crossbred sows compared to purebred nucleus sows which are voluntarily culled earlier to increase the speed of genetic progress. Sows with low soundness scores were at a higher risk of being culled than sows with high soundness scores.

Fernàndez de Sevilla *et al.* (2008) tested the effects of overall leg conformation score as well as six specific leg conformation defects on sow longevity in Duroc, Landrace, and Large White sows. Overall leg conformation score influenced sow longevity in all three breeds ($P < 0.01$), with poorly conformed sows being more likely to be culled than well-conformed sows. Plantigrade sows had decreased longevity in all three breeds ($P < 0.001$, Duroc; $P < 0.05$, Large White; $P < 0.1$, Landrace). Abnormal hoof growth affected survival in Duroc ($P < 0.001$) and Landrace ($P < 0.01$) sows. The survival of Duroc sows was also influenced by the presence of splayed feet ($P < 0.05$) and bumps and injuries ($P < 0.001$). Large White sows were affected by straight pastern ($P < 0.01$). Selection for improved leg conformation could indirectly lead to increased longevity in sows.

Relationship between Reproductive Traits and Carcass and Meat Quality Traits

Holm *et al.* (2004) investigated genetic correlations in Norwegian Landrace females between reproductive traits and lean meat content and bacon side quality. Most correlations were small and insignificant. However, there was a moderate unfavorable genetic correlation between lean meat content and NBA-P2. Sobczyńska *et al.* (2013) estimated lean meat percentage (LMP) using BF and *longissimus* muscle depth measurements taken during development adjusted to 180 days of age. A moderate and unfavorable genetic correlation existed between LMP and several longevity traits, including length of life (defined as days from birth to culling) and length of productive life (defined as days between first and last farrowing), number of litters, and lifetime pig production in Polish Large White sows. In Polish Landrace sows, while all correlations were unfavorable, only the correlation between LMP and length of life was significant.

Serenius *et al.* (2004) estimated genetic correlations between prolificacy and carcass and meat quality traits in Finnish Landrace and Large White pigs. Genetic correlations between prolificacy and carcass traits were largely unfavorable. Significant correlations existed between age at first farrowing (AFF) and lean percent (0.19 in Landrace and 0.27 in Large White) and fat percent (-0.26 in Landrace and -0.18 in Large White). While not as substantial, unfavorable correlations existed between both carcass traits and SB and first farrowing interval (FFI) in both breeds. There were no clear associations between prolificacy and meat quality traits. However, meat quality tended to be unfavorably correlated with TNB and favorably correlated with AFF and piglet mortality during suckling.

COMPONENTS OF SOW REPRODUCTIVE LONGEVITY

Reproductive longevity is a composite trait with many components, such as conception rate and WEI. Environmental factors play a large role in reproductive longevity and its component traits. Reproductive longevity is expressed throughout the life of a sow and is dependent on the physiological potential of the sow to resume ovarian cyclicity, rebreed, and farrow following successful parities (Tart *et al.*, 2013).

Drickamer *et al.* (1997) tested the effects of several factors on sow conception rates in Duroc-Hampshire-Yorkshire crossbred sows. Season during which breeding occurred did not have a significant effect on reproductive success, but mode of insemination did, with natural matings more likely to be successful than artificial insemination. Females who were only mated once were less likely to be successful, and sows mated two or more times were more likely to achieve reproductive success. There was no difference in success rate when females were inseminated twice versus three or more times. A sow's birth litter size did not have a significant impact on reproductive success; however, the sow's birth litter sex ratio did influence reproductive success. Sows originating from litters in which 1/3 or less of the litter was male had greater probability of successful inseminations and fewer failures than expected. Sows originating from litters in which 1/3 to 2/3 of the litter was male did not have more successful inseminations or failures than expected. Finally, sows originating from litters in which 2/3 or more of the litter was male had significantly fewer successful inseminations and more failures than expected.

Knauer *et al.* (2010b) identified several factors associated with STAY4 in six commercial maternal lines. Age at first farrowing was significantly associated with STAY4 in all six genetic

lines. Age at puberty became significant when AFF was not included in the model, suggesting that increased AP and AFF are early indicators of future reproductive problems and reduced sow longevity. Increased lactation feed intake was also significantly associated with STAY4 in all lines as greater feed intake during lactation increased probability to reach parity 4. Number born alive in a sow's final parity was positively associated with STAY4 in three lines. Lactation lengths of less than 11 days negatively affected STAY4 in five of the lines, whereas lactation lengths of 14 days or more were equally favorable for longevity. Number weaned, litter weaning weight, pre-farrow BF, and BF loss during lactation did not significantly influence STAY4. In 2006, Serenius *et al.* studied the same genetic lines to determine the effects of similar traits on culling risk. They found that AFF was positively associated with culling risk in one line. Litter size at first farrowing was associated with culling risk in two lines, with sows producing intermediately-sized litters having the lowest risk of being culled. Increased feed intake resulted in a decreased culling risk in all lines but one, and greater BF loss during lactation increased culling risk in all lines except two.

Hoge and Bates (2011) studied effects of several developmental factors on sow longevity in commercial Yorkshire sows, using six different traits that define of sow longevity. Across all traits, AFF, litter size at first and last farrowing, and SB in the first litter were significantly ($P < 0.0001$) associated with longevity. Sows that farrowed younger likely reached puberty at a younger age so they were mated at a younger age. These sows with a younger AFF exhibited improved reproductive performance. Larger litter sizes at first and last parity and fewer stillborn pigs and heaver litters at 21 days of lactation in parity 1 led to a decreased risk of being culled.

ECONOMIC IMPORTANCE OF REPODUCTIVE TRAITS

Optimization of reproductive performance is important as it can have large impacts on an operation's economic returns. Stalder *et al.* (2003) determined that a sow must remain in the herd for three parities in order to reach a positive net present value and recover the initial replacement gilt investment in a breed-to-wean operation based on 1996-2000 data and prices. Increasing pigs born alive per litter will increase net present value of sows. However, at 10.1 pigs born alive per litter (which was the average used in the analysis), a decrease of 0.5 pigs born alive would require four parities to reach a positive net present value, whereas an increase of four pigs born alive per litter would be required to reach a positive net present value by parity 2 if all other variables were held constant. Rodriguez-Zas *et al.* (2003) also stressed the importance of sow reproductive longevity on profitability. They studied data collected between 1995 and 2001 from 32 herds in central Illinois that were composed of eight major genetic lines. At a net income of $50 per parity, two parities were required to cover the replacement gilt cost. However, when a discount rate of 10% was applied to account for inflation and interest rates and risk, 3.68 parities were required recover replacement gilt costs. Genetic line chosen made a difference in profitability as the difference in average sow net present value between the best and worst line for longevity was $52.39.

Furthermore, Lamberson and Safranski (2000) performed simulations to optimize estrus detection and insemination schedules in order to optimize profitability by identification of a schedule that will result in high conception rate and litter size while minimizing labor and semen costs. Schedules with three or four inseminations resulted in greater reproductive performance than those with fewer inseminations when the simulation results of 500 herds of 100 sows each

were averaged. Though there was considerable variability in economic returns among herds within a particular schedule, the most profitable schedule was one with three inseminations at 12, 24, and 36 hours after first detection of estrus combined with estrus detection performed twice daily. A four insemination schedule with inseminations at 0, 12, 24, and 36 hours following first detection of estrus with estrus detection only performed once daily was a close second. Schedules with only one or two inseminations were less costly, but not as profitable due to decreased reproductive performance.

While genetic progress leads to increased economic returns, it is important to find the proper balance between progress and economics, as the system that leads to the highest genetic gain may not be the most profitable. This point was illustrated by Faust *et al.* (1992 and 1993). A computer model was developed to simulate pigs in a three-tiered breeding system being selected on an index of NBA, ADG, and BF for ten years. With industry standard voluntary and involuntary culling as well as death rates, selection was effective and profitable. Yearly increase in progeny breeding values for NBA were 0.126, 0.084, and 0.061 in the nucleus, multiplier, and commercial herds, respectively. Phenotypic changes in number of pigs weaned were 0.093, 0.099, and 0.059 in the nucleus, multiplier, and commercial herds, respectively. Index selection increased returns per finished pig by $1.35, $0.66, and $0.85 in the respective herds (Faust *et al.*, 1992). Faust *et al.* (1993) went on to use the same simulation model to compare systems with various combinations of culling after a maximum of one, five, and ten parities in each tier of the breeding structure. In the commercial herd, genetic progress for NBA and NW as well as efficiency traits was maximized when a maximum of one parity was allowed in all three tiers. Phenotypic means were greatest for NBA and NW in the system where maximum number of parities in the nucleus, multiplier, and commercial herds were one, one, and five or five, one, and five, respectively. While the system that allowed only one parity

across all three tiers had the highest income due to the most genetic progress, it also had the highest costs, resulting in the lowest returns of any system studied. Systems that allowed ten parities in the commercial herd and one parity in the nucleus herd were the most profitable. These systems could pay more for replacement gilts of higher genetic merit as they had lower replacement rates, whereas a system with a higher replacement rate may have to settle for a cheaper, lower genetic merit alternative. While the systems with the largest genetic change were not the most profitable, genetic progress is still very important as a comparable system with no genetic progress would always be less profitable (Faust *et al.*, 1993).

SELECTIVE SWEEPS

A selective sweep, also referred to as a signature of selection, occurs when a favorable allele increases in frequency within a population in response to natural or artificial selection pressure, simultaneously increasing the frequency of linked neutral alleles located nearby (Smith and Haigh, 1974; Nielsen, 2005). The favorable allele can either be a pre-existing allele that was previously rare or a new mutation that has arisen in the population (Hermisson and Pennings, 2005; Messer and Petrov, 2013). There are two types of selection sweeps, hard and soft. Hard sweeps occur when a single new mutation is the target of selection. As the mutation allele is driven to fixation, the ancestral variation of linked polymorphic loci decreases as the linked alleles "hitch-hike" with the selected allele and are also driven toward fixation. Recombination between "hitch-hiking" alleles and the selected allele is the only way to preserve ancestral variation. In contrast, a soft sweep occurs when the selected allele evolved independently multiple times. Ancestral variation may be retained as haplotype structure in the population under a soft sweep scenario as the selected allele may not be in complete linkage disequilibrium with surrounding variation; thus hard sweeps produce a signature of selection that is more easily identified than that which is generated by a soft sweep (Hermisson and Pennings, 2005).

There are many various statistics commonly used to aid in the identification of selective sweeps, including F_{ST} and extended haplotype homozygousity (EHH). F_{ST} is a measure of genetic differentiation between, rather than within, populations or subpopulations. Selection may favor different allelic variants in one population as compared to another, and the regions where this occurs will harbor more extreme allelic frequency differences than neutral regions or regions where selection favors the same allele in both populations (Qanbari *et al.*, 2011). It was first

described by Wright (1951) as “the correlation between random gametes, drawn from the same subpopulation, relative to the total.” This definition has since been interpreted in multiple ways, leading to different methods of calculating statistics all referred to as F_{ST}. The three general methods of F_{ST} calculation are method-of-moments, maximum-likelihood, and Bayesian methods. The method-of-moments approach analyzes variation in allelic frequencies between multiple populations to determine whether or not a difference exists and the extent of the difference. The maximum-likelihood approach mandates that a probability distribution from which samples originate be defined prior to calculation. The Bayesian approach also requires a probability distribution, but the entire outcome is not based on a distribution defined prior to sampling. Estimates are calculated utilizing a Markov chain Monte Carlo (MCMC) method (Porto-Neto *et al.*, 2013).

EHH was defined by Sabeti *et al.* (2002) as “the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent for the entire interval from the core region to a certain point.” The EHH statistic is used to identify regions targeted by artificial selection where allelic frequencies have increased faster than expected via drift or natural selection. The speed of allelic frequency increase is measured by the length of the surrounding conserved haplotype because recombination will shorten haplotype blocks over time. If an allele takes longer to reach fixation, the surrounding linked haplotype will be shorter than if the allele was rapidly driven to fixation by artificial selection (Gärke *et al.*, 2014). Relative extended haplotype homozygosity (REHH) corrects for differences in local recombination rates by comparing the rate of EHH decay on the core haplotype being tested to the combined rate of EHH decay on other core haplotypes in the region (Sabeti *et al.*, 2002). F_{ST} is better able to detect selective sweeps when variation is fixed, whereas EHH can detect selective sweeps in regions that are still segregating (Qanbari *et al.*, 2011). While EHH requires accurate

chromosome phasing and ancestral allele identification, which is not always easy to obtain (Porto-Neto *et al.*, 2013), it is less sensitive to ascertainment bias than other methods (Zhang *et al.*, 2012), thus making it a good approach when using SNP data.

Identifying selective sweeps provides a valuable link between phenotype and genotype and has the potential to improve breeding programs by identifying regions and then specific favorable genotypes to target via genomic selection (Rothammer *et al.*, 2013). The selective sweep approach is also advantageous as it allows the detection of favorable genotypes that have reached fixation that would not be detectable with traditional "forward" methods, which seek to identify functional genes and mutations beginning with a phenotype in contrast to beginning with a signature of selection (Ramey *et al.*, 2013). Advanced technologies, including high-throughput genotyping methods, have made the detection of selective sweeps possible (Rothammer *et al.*, 2013). However, random drift may produce genomic regions with the same characteristics as a selective sweep that are indistinguishable from true selective sweeps (Ramey *et al.*, 2013). Nonetheless, many studies have been conducted in recent years with the objective of identifying selective sweeps that have occurred as a result of domestication, breed differentiation, or artificial selection within a breed or population in a wide variety of species ranging from cattle to dogs.

Many selection sweeps have been identified in cattle; some overlap between breeds, though most were breed-specific. Some of these regions harbor potential candidate genes and regulatory regions, while others did not have any annotated genes implicated in traits that may have been selected for during domestication and breed formation (Qanbari *et al.*, 2011; Ramey *et al.*, 2013; Rothamer *et al.*, 2013).

Phua *et al.* (2014) used a selective sweep approach to compare two lines of Romney sheep divergently selected for facial eczema resistance and susceptibility in order to identify loci involved in the development of the disease. Two different statistics were calculated, and the three most significant markers identified were the same between the two approaches. Additional significant regions were identified when windows were analyzed rather than markers. In total, eight regions appeared to have undergone a selective sweep. These regions were searched for positional candidate genes, but none were found. Little overlap exists between the results of this study and those of previous studies attempting to locate the genes responsible for facial eczema, suggesting that many loci with small effects may contribute to the development of this disease.

A QTL was identified by Van Laere *et al.* (2003) that explained 15-30% of phenotypic variation in muscle mass and 10-20% of variation in back-fat thickness in pigs. It was predicted that a new allele had arisen that promoted muscle development in the QTL region. A 250-kb region was identified on SSC9 in which all heavily muscled individuals shared a haplotype. It was assumed that the QTL was located in this region, which contains the genes *INS* and *IGF2*. Sequencing revealed 258 polymorphisms and two divergent haplotype clusters. One polymorphism, located in *IGF2* intron 3 in an evolutionarily conserved CpG island, was identified where one allele was present in all heavily muscled individuals and the other allele was present in all “wild-type” individuals. The mutation was validated by genotyping additional individuals and analyzing gene expression. It was not present in individuals from lines that were not selected for lean growth but was present in lines that were. There are significant differences in *IGF2* expression between genotypes in muscle tissues, but not other tissues.

Rubin *et al.* (2012) compared domestic pig and wild boar populations in order to identify selective sweeps displaying considerable differences in allelic frequency. The strongest signature

of selection was located in a QTL region for number of vertebrae in pigs on SSC1 and contained the gene *NR6A1* (*Nuclear Receptor 6 A1*), which harbors a mutation thought to be the cause of vertebrae number variation. Two other loci showing substantial evidence of a selective sweep overlapped QTLs for body length, which explained 18.4% of the residual variance in body length and together acted additively to produce a 5.3 cm difference in body length between opposite homozygotes. The two loci are located on SSC4 and SSC8 and contain *PLAG1* (*pleomorphic adenoma gene 1*) and *LCORL* (*ligand dependent nuclear receptor corepressor-like*) as candidate genes. Genotyping of these regions in additional individuals from a wide range of European domestic pig populations, European wild boars, and Asian domestic pig populations revealed strong signatures of selection at all three loci in all European domestic pigs used for meat production. Another substantial selective sweep was located on SSC13 and contained the gene *OSTN* (*Osteocrin*), which has differential expression between muscle fiber types; therefore, this sweep may be a result of selection for altered body composition and skeletal development. Sequencing was used to identify polymorphisms that displayed allelic frequency differences between domestic pigs and wild boars; however, very little overlap existed between these polymorphisms and the previously identified sweep regions. While not in a sweep region, a series of duplications at the *KIT* locus were identified that are causative of white, patch, or belt coloration. A challenge in identifying selective sweeps arises due to the recent divergence of domestic pigs from wild boars as their genomes are still very similar, and only loci under very strong selection display sequence differences that are substantial enough to be identified as a selective sweep.

Li *et al.* (2014) compared Large White and Tongcheng pigs, identifying 34 and 25 regions, respectively, in these breeds that appear to have undergone a selective sweep. Most of the candidate genes identified within these regions were involved in growth, reproduction, and

immune response, with a stronger focus on immune response within the Tongcheng population. There was not much overlap in selected regions between these two breeds as they are genetically very different from each other, indicated by an average F_{ST} value of 0.254.

Gärke *et al.* (2014) searched for signatures of selection in the Göttingen Minipig (GMP), a breed developed for medical research and toxicology at the University of Göttingen. It is a combination of three different breeds, each with desired characteristics; the Vietnamese potbellied pig was included for high fertility, the Minnesota Minipig was used for small size, and the German Landrace was introduced for white color. The resulting GMP is a white, dwarf animal with all body parts reduced in size. Regions of the genome were identified that contained different proportions of breed composition than what was expected based on pedigree information. Sixty significant signals were detected, with a large proportion occurring on chromosomes 1, 11, and 15. These regions and the surrounding area (1-Mb up- and down-stream of the signal) were searched for candidate genes, and two genes with functions obviously related to the selection goals of the GMP were identified. The *discoidin domain receptor tyrosine kinase 2* (*DDR2*) gene, located on chromosome 4, is implicated in body size in mice and could be partially responsible for the small body size of the GMP. The *prolactin receptor* (*PRLR*) gene, located on chromosome 16, is associated with number of piglets born alive and number of teats. The REHH statistic was employed to identify regions with core haplotypes in strong linkage disequilibrium. Clusters of signals were observed on several chromosomes, with the highest signals located on chromosomes 3, 5, 7, 9, and 14. Several candidate genes were located within these significant regions. The *suppressor of cytokine signaling 2* (*SOCS2*) gene, located on chromosome 5, negatively regulates growth hormone and insulin-like growth factor-1 (IGF-1) and may contribute to the reduced size of the GMP. This gene is also located in a region that deviates significantly from the expected breed composition,

with an over-representation of the Minnesota Minipig and under-representation of German Landrace. The *thioredoxin* (*TXN*) gene, located on chromosome 1, may have an effect on growth-related traits in pigs, and the *BMP6* gene, located on chromosome 7, is a member of the bone morphogenic protein family, which plays a role in bone growth as well as ovarian function and follicular development. The *GAB2* gene, located on chromosome 9, is a member of a family (GRB2-associated binding protein gene family) that is correlated with various cytokines and growth factors, implicating a potential influence on body size. Contrary to expectations, there were not any candidate genes identified in these regions that influenced coat color despite strong selection for this trait. A chi-square test was performed for the deviation in breed composition, yielding additional significant regions containing candidate genes. The *growth factor receptor-bound protein 10* (*GRB10*) gene, located on chromosome 9, has a strong influence on animal growth. The *mechanistic target of rapamycine* (*serine/threonine kinase*; *MTOR*) gene, located on chromosome 6, is involved in a pathway that regulates growth factor signaling, and its inhibition results in changes in GRB10 abundance.

Makvandi-Nejad *et al.* (2012) performed a genome-wide association study that identified four loci that explain most of the size variation in horses. Three of these loci contained genes previously known to affect size in other species, *ligand dependent nuclear receptor corepressor-like* (*LCORL*), *HMGA2*, and *zinc finger and AT hook domain containing* (*ZFAT*). The locus containing *LCORL* shows indications of a selective sweep in large breeds as haplotype diversity is very low.

Petersen *et al.* (2013) located signatures of selection in 33 horse breeds, with a focus on regions that appear to be involved in coat color, performance, gait, and size. A shared haplotype was present in chestnut colored breeds near the *MC1R* locus, known for involvement in coat color. A selective sweep was also identified in a dun colored breed in a region previously

associated with dun coloration. A haplotype was identified in the American Paint Horse and Quarter Horse, breeds known for sprinting ability, which includes the myostatin gene (*MSTN*). Variants in this gene were associated with altered muscle fiber type proportions that are favorable for sprinting. A region on ECA23 was highly significant in four breeds which all possess alternative gaits, and a sweep on ECA11 was significant in draft and miniature breeds, suggesting involvement in size determination.

Rubin *et al.* (2010) identified selective sweeps that occurred during chicken domestication and specialization into broilers and layers. Four different broiler lines, including two lines divergently selected for high and low growth, four different layer lines, and red jungle foul, the major wild ancestor of the domestic chicken, were compared. The most significant sweep identified when all domestic lines were compared to red jungle fowl overlapped the gene encoding thyroid stimulating hormone receptor (*TSHR*), known for involvement in metabolic regulation and reproduction. This 40-kb sweep region displayed nearly complete fixation in all domestic lines. Additional domestic birds (n = 271) from 36 geographically diverse populations were genotyped in this region, and all but seven birds were homozygous for the sweep haplotype. A non-conservative amino acid substitution was identified that is predicted to influence ligand interaction in the translated protein. Several selective sweeps were identified in the broiler populations in regions that influence muscle growth, including regions containing the genes coding for insulin-like growth factor 1 (*IGF1*), pro-melanin-concentrating hormone (*PMCH*), and *TBC1D1*. A deletion was identified in the *growth hormone receptor* (*GHR*) gene, which has been found to cause sex-linked dwarfism and has been used to decrease growth and feed consumption in some broiler parental lines. Another deletion was identified in *SH3 domain containing ring finger 2* (*SH3RF2*), located in a QTL for body weight. It was fixed in the high growth line and occurs at a low frequency in the low growth line. An association analysis on

individuals from an intercross of the two selection lines confirmed an association between the presence of the deletion and increased growth.

Selective sweeps were identified in two broiler lines divergently selected for abdominal fat deposition after eleven years of selection (Zhang *et al.*, 2012). EHH tests were calculated on the core regions identified, and 51 and 57 core regions were found to be significant ($P < 0.01$) in the lean and fat lines, respectively. A large proportion of these selective signals were located on chromosomes 1, 2, 3, and 4. Ten candidate genes, including *RB1* (retinobstoma 1), *BBS7* (Bardet-Biedl syndrome 7), *MAOA* (monoamine oxidase A), *MAOB* (monoamine oxidase B), *EHBP1* (EH domain binding protein 1), *LRP2BP* (LRP2 binding protein), *LRP1B* (low-density lipoprotein receptor-related protein 1B), *MYO7A* (myosin VIIA), *MYO9A* (myosin IXA), and *PRPSAP1* (phosphoribosyl pyrophosphate synthetase-associated protein 1), were identified that are known to have functions which could affect fatness. Seven of these ten candidate genes were located within published QTL for abdominal fat content in chickens.

Pollinger *et al.* (2005) used simulations to test the efficacy of using various homozygosity statistics as well as the F_{ST} statistic to locate selective sweeps in the dog genome. Through simulation, they also proved that with moderately spaced, highly variable markers, the power to detect selective sweeps is high while Type I error rate is low. They were then able to locate selective sweeps in the Large Munsterlander and Dachshund in the genomic regions that were previously found to be responsible for black coat color and achodroplasia, respectively.

Sutter *et al.* (2007) identified two QTL for body size on chromosome 15 in Portuguese water dogs, a breed known for size variation. They then sequenced the regions and performed an association analysis between 116 SNPs and skeletal size, which identified a single peak near the insulin-like growth factor 1 (*IGF1*) gene. Haplotype analysis in this region revealed that 15%

of the variation in skeletal size is explained by *IGF1* haplotype. The area was identified as a selective sweep as it displayed a substantial decrease in heterozygosity and increase in differentiation between large and small breeds. Association analysis between the SNPs in the sweep region and skeletal size resulted in a significant 84-kb region containing 25 SNPs. Only three haplotypes were identified; one was shared by small breeds, and two were found in large breeds. This selective sweep probably occurred early in the domestication process as the same haplotypes, which are relatively short due to recombination over time, are shared by distantly related and reproductively isolated breeds.

Akey *et al.* (2010) identified 155 selection sweeps in ten phenotypically diverse dog breeds, of which 103 were breed-specific or only found in a couple breeds. Many of these regions contained candidate genes for conspicuous phenotypes, including size, coat color and texture, behavior, skeletal morphology, and physiology. For example, regions harboring the *IGF1* gene, which is implicated in miniature size, and the *HMGA2* gene, which results in a pygmy phenotype in mice, were identified in toy and small breeds, respectively. The region containing *HAS2*, a hyaluronic acid synthase, was identified as a selection sweep in the Shar-Pei breed, known for excessive skin wrinkling that is associated with high levels of mucin and hyaluronic acid. Further analysis of this gene revealed a two bp deletion that was significantly associated with the degree of wrinkled skin in Shar-Pei and was not present in any other breed. Olsson *et al.* (2011) also scanned the Shar-Pei genome for selective sweeps in order to locate candidate genes for the hyaluronanosis phenotype observed in the Shar-Pei breed. The strongest signal of reduced heterozygosity corresponded to the region containing *HAS2*, and GWAS identified SNPs in this region to be significantly associated with familial Shar-Pei fever, an autoinflammatory disease unique to the Shar-Pei breed that resembles some human hereditary periodic fever syndromes. Sequencing of the region revealed two overlapping duplications, and an association

was observed between copy number and fever symptoms, suggesting that the duplication is a causative mutation of both hyaluronanosis and Shar-Pei fever.

Quilez *et al.* (2011) scanned the Boxer genome and genomes of other dog breeds for selective sweeps. A sweep was identified on chromosome 1 in the Boxer and other brachycephalic breeds that has been previously associated with brachycephaly. In addition, a novel sweep was identified on chromosome 26 in the Boxer that was also present, though shorter in length, in similar breeds that share an evolutionary history. The phenotypic traits that correlate to this sweep remain unknown.

Selection sweeps were identified in the Cornish Rex breed and eleven other phenotypically diverse domestic cat breeds in order to identify the locus responsible for the curly rexoid hair texture (Gandolfi *et al.*, 2013). A region on chromosome A1 demonstrated evidence of a strong selective sweep detected via three different methods in the Cornish Rex but not in any of the other populations tested. This region also overlapped a homozygous block identified in the Cornish Rex that was not present in any other population studied. The region was searched for candidate genes, and *lysophosphatidic acid receptor 6* (*LPAR6*), known for involvement in hair growth and texture, was identified. Sequencing of this gene revealed a 4 bp deletion that causes a frameshift mutation and introduces a premature stop codon. This deletion was homozygous in all Cornish Rex individuals and was present in the German Rex population. However, it was not present in any other breed or population studied, including other rex breeds, indicating that it is only responsible for the curly coat phenotype in Cornish and German Rex breeds.

Werzner *et al.* (2013) analyzed a genomic region on *D. melanogaster* chromosome X that was previously identified as a selective sweep, presumably occurring during the evolution

of the European population from the African population. Sequencing of this region, containing the *Flo-2* gene, revealed a pattern typical of a selective sweep; there was a 10-kb region of reduced genetic variation with the surrounding regions gradually increasing to neutral levels in the European population. This region was also identified as one with a statistically significant deviation from neutral expectations. Eleven nucleotide substitutions that are differentially fixed between the African and European populations were identified. Upon comparison to the common ancestor of both populations, eight substitutions in the European population differed from the ancestral allele, indicating that they may have arisen after the divergence of the European population. The overrepresentation of derived alleles in this segment is statistically significant.

Whole genome sequencing was used by Liu *et al.* (2014) to estimate the time of divergence between polar bears and brown bears, infer their demographic history, and detect genes under positive selection in polar bears to better understand how polar bears were able to adapt to the harsh Arctic environment. Divergence time was estimated to be only 479-343 thousand years ago, and evidence of gene flow from polar bears into North American brown bears immediately following divergence was present, though gene flow has not occurred recently. Adaptation to the Arctic occurred prior to 110 thousand years ago in less than 20,500 generations. Selective sweeps and candidate genes were identified in the polar bear genome. There was an overrepresentation of genes associated with adipose tissue development and cardiovascular function within the sweep regions, indicating adaptations that allow polar bears to cope with a high fat diet, sizeable adipose tissue deposits, and very high levels of cholesterol in blood plasma. Two candidate genes associated with pigmentation were also found in selective sweep regions that may be involved in producing the characteristic white coloration of polar bears.

LITERATURE CITED

Abell, C.E., Jones, G.F., Stalder, K.J., P.A.S., and Johnson, A.K. 2010. Using the Genetic Lag Value to Determine the Optimal Maximum Parity for Culling in Commercial Swine Breeding Herds. The Professional Animal Scientist. 26:404-411.

Abell, C.E., Mabry, J.W., Dekkers, J.C.M., and Stalder, K.J. 2013. Relationship between litters per sow per year sire breeding values and sire progeny means for farrowing rate, removal parity and lifetime born alive. Journal of Animal Breeding and Genetics. 130:64-71.

Akey, J.M., Ruhe, A.L., Akey, D.T., Wong, A.K., Connelly, C.F., Madeoy, J., Nicholas, T.J., and Neff, M.W. 2010. Tracking footprints of artificial selection in the dog genome. PNAS 107(3):1160-1165.

Bazer, F.W. and Johnson, G.A. 2014. Pig blastocyst-uterine interactions. Differentiation. 87:52-65.

Bennett, G.L. and Leymaster, K.A. 1990b. Genetic implications of a simulation model of litter size in swine based on ovulation rate, potential embryonic viability and uterine capacity: II. Simulated selection. Journal of Animal Science. 68:980-986.

Bidanel, J.P. 2011. Biology and genetics of reproduction. In *The Genetics of the Pig.* 2nd Edition. Edited by Rothschild, M.F. and Ruvinsky, A. Wallingford, UK: CAB International; 218-241.

Bjerre, D., Mark, T., Sørensen, P., Proschowsky, H.F., Vernersen, A., Jørgensen, C.B., and Fredholm, M. 2010. Investigation of candidate regions influencing litter size in Danish Landrace sows. Journal of Animal Science. 88:1603-1609.

Bolet, G., Bidanel, J.P. and Ollivier, L. 2001. Selection for litter size in pigs. II. Efficiency of closed and open selection lines. Genet. Sel. Evol. 33:515-528.

Canario, L., Cantoni, E., Le Bihan, E., Caritez, J.C., Billon, Y., Bidanel, J.P., and Foulley, J.L. 2006b. Between-breed variability of stillbirth and its elationship with sow and piglet characteristics. Journal of Animal Science. 84:3185-3196.

Cassady, J.P., Young, L.D., and Leymaster, K.A. 2002. Heterosis and recombination effects on pig reproductive traits. Journal of Animal Science. 80:2303-2315.

Chen, P., Baas, T.J., Mabry, J.W., and Koehler, K.J. 2003. Genetic correlations between lean growth and litter traits in U.S. Yorkshire, Duroc, Hampshire, and Landrace pigs. Journal of Animal Science. 81:1700-1705.

Dekkers, J.C.M., Mathur, P.K, and Know, E.F. 2011. Genetic improvement of the pig. In *The Genetics of the Pig.* 2nd Edition. Edited by Rothschild MF and Ruvinsky A. Wallingford, UK: CAB International; 390-425.

Distl, O. 2007. Mechanisms of regulation of litter size in pigs on the genome level. Reprod. Dom. Anim. 42(Suppl. 2):10-16.

Drickamer, L.C., Arthur, R.D., Rosenthal, T.L. 1997. Conception failure in swine: importance of the sex ratio of a female's birth litter and tests of other factors. Journal of Animal Science. 75:2192-2196.

Faust, M.A., Tess, M.W., and Robison, O.W. 1992. A bioeconomic simulation model for a hierarchical swine breeding structure. Journal of Animal Science. 70:1760-1774.

Faust, M.A., Robison, O.W., and Tess, M.W. 1993. Genetic and economic analyses of sow replacement rates in the commercial tier of a hierarchical swine breeding structure. Journal of Animal Science. 71:1400-1406.

Fernàndez de Sevilla, X., Fàbrega, E., Tibau, J., and Calellas, J. 2008. Effect of leg conformation on survivability of Duroc, Landrace, and Large White sows. Journal of Animal Science. 86:2392-2400.

Gandolfi, B., Alhaddad, H., Affolter, V.K., Brockman, J., Haggstrom, J., Joslin, S.E.K., Koehne, A.L., Mullikin, J.C., Outerbridge, C.A., Warren, W.C., and Lyons, L.A. 2013. To the root of the curl: a signature of a recent selective sweep identifies a mutation that defines the Cornish Rex cat breed. PLOS ONE 8(6):e67105.

Gärke, C., Ytournel, F., Sharifi, A.R., Pimentel, E.C.G., Ludwig, A., and Simianer, H. 2014. Footprints of recent selection and variability in breed composition in the Göttingen Minipig genome. Stichting International Foundation for Animal Genetics 45:381-391.

Hamann, H., Drögemüller, C., Krieter, J., Presuhn, U., Wallenburg, J., Distl, O. 2000. Genetic markers for litter size in German pig breeds. In: 51st Annual Meeting of the European Association of Animal Production, The Hague, 21-24 August 2000.

Hermisson, J., and Pennings, P.S. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335-2352.

Hoge, M.D. and Bates, R.O. 2011. Developmental factors that influence sow longevity. Journal of Animal Science. 89:1238-1245.

Holl, J.W. and Robison, O.W. 2003. Results from nine generations of selection for increased litter size in swine. Journal of Animal Science. 81:624-629.

Holm, B., Bakken, M., Klemetsdal, G., and Vangen, O. 2004. Genetic correlations between reproduction and production traits in swine. Journal of Animal Science. 82:3458-3464.

Hsu, W.L. 2011. Analysis of long-term selection (28 generations) for reproduction, growth, and carcass traits in swine. *PhD dissertation.* University of Nebraska – Lincoln, Animal Science Department.

Jamann, T.M., Balint-Kurti, P.J., and Holland, J.B. 2015. QTL Mapping Using High-Throughput Sequencing. In *Plant Functional Genomics: Methods and Protocols.* Methods in Molecular Biology, vol. 1284. Edited by Alonso, J.M. and Stepanova, A.N. New York: Springer Science+Business Media; 257-285.

Johnson, R.K., and Omtvedt, I.T. 1975. Maternal heterosis in swine: reproductive performance and dam productivity. Journal of Animal Science. 40(1):29-37.

Johnson, R.K., Nielsen, M.K., and Casey, D.S. 1999. Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. Journal of Animal Science. 77:541-557.

Knauer, M.T., Cassady, J.P., Newcom, D.W., and See, M.T. 2010a. Estimates of variance components for genetic correlations among swine estrus traits. Journal of Animal Science. 88:2913-1919.

Knauer, M., Stalder, K.J., Serenius, T., Baas, T.J., Berger, P.J., Karriker, L., Goodwin, R.N., Johnson, R.K., Mabry, J.W., Miller, R.K., Robison, O.W., and Tokach, M.D. 2010b. Factors associated with sow stayability in 6 genotypes. Journal of Animal Science. 88:3486-3492.

Knauer, M.T., Cassady, J.P., Newcom, D.W., and See, M.T. 2011. Phenotypic and genetic correlations between gilt estrus, puberty, growth, composition, and structural conformation traits with first-litter reproductive measures. Journal of Animal Science. 89:935-942.

Kuhlers, D.L., Jungst, S.B., and Little, J.A. 1994. An experimental comparison of equivalent terminal and rotational crossbreeding systems in swine: sow and litter performance. Journal of Animal Science. 72:584-590.

Lamberson, W.R., Johnson, R.K., Zimmerman, D.R., and Long, T.E. 1991. Direct responses to selection for increased litter size, decreased age at puberty, or random selection following selection for ovulation rate in swine. Journal of Animal Science. 69:3129-3143.

Lamberson, W.R. and Safranski, T.J. 2000. A model for economic comparison of swine insemination programs. Theriogenology. 54:799-807.

Li, N., Zhoa, Y.F., Xiao, L., Zang, F.J., Chen Y.Z., Dai, R.J., Zang, J.S., Shen, S.Q., Chen, Y.F., Wu C.X. 1998. Candidate gene analysis for identification of genetic loci controlling litter size in swine. In: Proceedings of the Sixth World Congress on Genetics Applied to Livestock Production, Armidale, vol. 26, pp. 183-186.

Li, X., Yang, S., Tang, Z., Li, K., Rothschild, M.F., Liu, B., and Fan, B. 2014. Genome-wide scans to detect positive selection in Large White and Tongcheng pigs. Animal Genetics 45:329-339.

Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., Zhou, L., Korneliussen, T.S., Somel, M., Babbitt, C., Wray, G., Li, J., He, W., Wang, Z., Fu, W., Xiang, X., Morgan, C.C., Doherty, A., O'Connell, M.J., McInerney, J.O., Born, E.W., Dalén, L., Dietz, R., Orlando, L., Sonne, C., Zhang, G., Nielsen, R., Willerslev, E., and Wang, J. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptations in polar bears. Cell. 157:785-794.

Lillehammer, M., Meuwissen, T.H.E., and Sonesson, A.K. 2013. Genomic selection for two traits in a maternal pig breeding scheme. Journal of Animal Science. 91:3079-3087.

Makvandi-Nejad, S., Hoffman, G.E., Allen, J.J., Chu, E., Gu, E., Chandler, A.M., Loredo, A.I., Bellone, R.R., Mezey, J.G., Brooks, S.A., and Sutter, N.B. 2012. Four loci explain 83% of size variation in the horse. PLoS ONE. 7(7):e39929.

Marantidis, A., Papadopoulos, A.I., Michailidis, G, and Avdi, M. 2013. Association of BF gene polymorphism with litter size in a commercial pig cross population. Animal Reproduction Science. 141:75-79.

Messer, P.W., and Petrov, D.A. 2013. Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology & Evolution. 28:659-669.

Mészáros, G., Pálos, J., Ducrocq, V., and Sölkner, J. 2010. Heritability of longevity in Large White and Landrace sows using continuous time and grouped data models. Genetics Selection Evolution. 42:13-25.

Meuwissen, T.H.E., Hayes, B.J., and Goddard M.E. 2001. Prediction of total genetic value using genome-side dense marker maps. Genetics. 157:1819-1829.

Nielsen, B., Su, G., Lund, M.S., and Madsen, P. 2013. Selection for increased number of piglets at d 5 after farrowing has increased litter size and reduced piglet mortality. Journal of Animal Science. 91:2575-2582.

Nielsen, R. 2005. Molecular Signatures of Natural Selection. Annu. Rev. Genet. 39:197-218.

Noguera, J.L., Varona, L., Babot, D., and Estany, J. 2002. Multivariate analysis of litter size for multiple parities with production traits in pigs: II. Response to selection for litter size and correlated response to production traits. Journal of Animal Science. 80:2548-2555.

Nonneman, D., Lents, C., Rohrer, G., Rempel, L., Vallet, J. 2013. Genome-wide association with delayed puberty in swine. Animal Genetics. 45:130-132.

Olsson, M., Meadows, J.R.S., Truvé, K. Pielberg, G.R., Puppo, F., Mauceli, E. Quilez, J., Tonomura, N., Zanna, G., Docampo, M.J., Bassols, A., Avery, A.C., Karlsson, E.K., Thomas, A., Kastner, D.L., Bongcam-Rudloff, E., Webster, M.T., Sanchez, A., Hedhammar, Å., Remmers, E.F., Andersson, L., Ferrer, L., Tintle, L., and Lindblad-Toh, K. 2011. A novel unstable duplication upstream of *HAS2* predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. PLoS Genetics. 7(3):e1001332.

Onteru, S.K., Fan, B., Du, Z.Q., Garrick, D.J., Stalder, K.J., and Rothschild, M.F. 2011. A whole-genome association study for pig reproductive traits. Animal Genetics. 43:18-26.

Petersen, J.L., Mickelson, J.R., Rendahl, A.K., Valberg, S.J., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., Brama, P., Machado, A.C., Capomaccio, S., Cappelli, K., Cothran, E.G., Distl, O., Fox-Clipsham, L., Graves, K.T., Guérin, G., Haase, B., Hasegawa, T., Hemmann, K., Hill, E.W., Leeb, T., Lindgren, G., Lohi, G., Lopes, M.S., McGivney, B.A., Mikko, S., Orr, N., Penedo, M.C.T., Piercy, R.J., Raekallio, M., Reider, S., Røed, K.G., Swinburne, J., Tozaki, T., Vaudin, M., Wade, C.M., and McCue, M.E. 2013. Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS 9(1):e1003211.

Phua, S.H., Brauning, R., Baird, H.J., and Dodds, K.G. 2014. Identifying chromosomal selection-sweep regions in facial eczema selection-line animals using an ovine 50K-SNP array. Animal Genetics 45:240-247.

Pollinger, J.P., Bustamante, C.D., Fledel-Alon, A., Schmuntz, S., Gray, M.M., and Wayne, R.K. 2005. Selective sweep mapping of genes with large phenotypic effects. Genome Research. 15:1809-1819.

Pomp, D., Caetano, A.R., Bertani, G.R., Gladney, C.D., and Johnson, R.K. 2001. Applying function genomics research to the study of pig reproduction. Reproduction. Suppl. 58:27-292.

Porto-Neto, L.R., Lee, S.W., Lee, H.K., and Gondro, C. 2013. Detection of Signatures of Selection Using F_{ST}. In *Genome-Wide Association Studies and Genomic Prediction.* Methods in Molecular Biology, vol. 1019. Edited by Gondro, C., Van Der Werf, J., and Hayes, B. New York: Springer Science+Business Media; 423-435.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G., and Simianer, H. 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics 12:318-329.

Quilez, J., Short, A.D., Martínez, V., Kennedy, L.J., Ollier, W., Sanchez, A., Altet, L., and Francino, O. 2011. A selective sweep of >8 Mb on chromosome 26 in the Boxer genome. BMC Genomics 12:339-350.

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D., and Taylor, J.F. 2013. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics 14:382-400.

Rempel, L.A., Nonnemann, D.J., Wise, T.H. Erkens, T., Peelman, L.J., and Rohrer, G.A. 2010. Association analyses of candidate single nucleotide polymorphisms on reproductive traits in swine. Journal of Animal Science. 88:1-15.

Rodriguez-Zas, S.L., Southey, B.R., Knox, R.V., Connor, J.F., Lowe, J.F., and Roskamp, B.J. 2003. Bioeconomic evaluation of sow longevity and profitability. Journal of Animal Science. 81:2915-2922.

Rothammer, S., Seichter, D., Förster, M., and Medugorac, I. 2013. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC Genomics 14:908-924.

Rothschild, M.F., Jacobson, C., Vaske, D., Tuggle, C.K., Wang, L., Short, T., Eckardt, G., Sasaki, S., Vincent, A., McLaren, D., Southwood, O., van der Stehen, A., Mileham, A., and Plastow, G.S. 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proceedings of the National Academy of Sciences of the United States of America. 93:201-205.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Biochard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., and Andersson, L. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587-593.

Rubin, C.J., Megans, H.J., Barrio, A.M., Maqbool, K. Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A.M., and Andersson, L. 2012. Strong signatures of selection in the domestic pig genome. PNAS. 109(48):19529-19536.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., and Lander, E.S. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. 419:832-837.

Schneider, J.F, Rempel, L.A., Rohrer, G.A., and Brown-Brandl, T.M. 2011. Genetic parameter estimates among scale activity score and farrowing disposition with reproductive traits in swine. Journal of Animal Science. 89:3514-3521.

Schneider, J.F., Rempel, L.A., and Rohrer, G.A. 2012a. Genome-wide association study of swine farrowing traits. Part I: Genetic and genomic parameter estimates. Journal of Animal Science. 90:3353-3359.

Schneider, J.F., Rempel, L.A., Snelling, W.M., Wiedmann, R.T., Nonneman, D.J., and Rohrer, G.A. 2012b. Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data. Journal of Animal Science. 90:3360-3367

Serenius, T., Sevon-Aimónen, M.L., Kause, A., Mäntysaari, E.A., and Mäki-Tanila, A. 2004. Genetic associations of prolificacy with performance, carcass, meat quality, and leg conformation traits in Finnish Landrace and Large White pig populations.

Serenius, T. and Stalder, K.J. 2006. Selection for sow longevity. Journal of Animal Science. 84(E. Suppl.):E166-E171.

Serenius, T., Stalder, K.J., Baas, T.J., Mabry, J.W., Goodwin, R.N., Johnson, R.K., Robison, O.W., Tokach, M., and Miller, R.K. 2006. National Pork Producers Council Maternal Line National Genetic Evaluation Program: A comparison of sow longevity and trait associations with sow longevity. Journal of Animal Science. 84:2590-2595.

Serenius, T. and Stalder, K.J. 2007. Length of productive life of crossbred sows is affected by farm management, leg conformation, sow's own prolificacy, sow's origin parity and genetics. Animal. 1:745-720.

Short, T.H., Southwood, O.I., Devries, A.G., McLaren, D.G., Evans, G.J., Mileham, A.J., and Plastow, G.S. 1997b. Evidence of a new genetic marker for litter size in pgs. Journal of Animal Science. 75(Suppl 1):29.

Smith, J.M. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23(1):23-35.

Sobczyńska, M., Blicharski, T., and Tyra, M. 2013. Relationships between longevity, lifetime productivity, carcass traits, and conformation in Polish maternal pig breeds. Journal of Animal Breeding and Genetics. 130:361-371.

Soede, N.M., Langendijk, P., and Kemp, B. 2011. Reproductive cycles in pigs. Animal Reproduction Science. 124:251-258.

Spötter, A., and Distl, O. 2006. Genetic approaches to the improvement of fertility traits in the pig. The Veterinary Journal. 172:234-247.

Spötter, A., Drögemüller, C., Hamann, H., and Distl, O. 2005. Evidence of a new LIF associated genetic marker for litter size in a synthetic pig line. Journal of Animal Science. 83:2264-2270.

Stalder, K.J., Lacy, C., Cross, C.L., and Conatser, G.E. 2003. Financial impact of average parity of culled females in a breed-to-wean swine operation using replacement gilt net present value analysis. Journal of Swine Health and Production. 11(2):69-74.

Sutter, N.B., Bustamante, C.D., Chase, K., Gray, M.M., Zhao, K., Zhu, L., Padhukasahasram, B., Karlins, E., Davis, S., Jones, P.G., Quignon, P., Johnson, G.S., Parker, H.G., Fretwell, N., Mosher, D.S., Lawler, D.F., Satyaraj, E., Nordborg, M., Lark, G.K., Wayne, R.D., and Ostrander, E.A. 2007. A single *IGF1* allele is a major determinant of small size in dogs. Science 316:112-115.

Tao, H., Mei, S., Sun, X., Peng, X., Zhang, X., Ma, C., Wang, L., Hua, L., and Li, F. 2013. Associations of *TCF12, CTNNAL1,* and *WNT10B* gene polymorphisms with litter size in pigs. Animal Reproduction Science. 140:189-194.

Tart, J.K., Johnson, R.K., Bundy, J.W., Ferdinand, N.N., McKnite, A.M., Wood, J.R., Miller, P.S., Rothschild, M.F., Spangler, M.L., Garrick, D.J., Kachman, S.D., and Ciobanu, D.C. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Animal Genetics. 44:387-397.

Tomiyama, M., Kubo, S., Takagi, T., and Suzuki, K. 2011. Evaluation of genetic trends and determination of the optimal number of cumulative records of parity required in reproductive traits in a Large White pig population. Animal Science Journal. 82:621-626.

Tummaruk, P., Lundeheim, N., Einarsson, S., and Dalin, A.M. 2001. Effect of birth litter size, birth parity number, growth rate, backfat thickness, and age at first mating of gilts on their reproductive performance as sows. Animal Reproduction Science. 66:225-237.

Uimari, P., Sironen, A., and Sevón-Aimonen, M.L. 2011. Whole-genome SNP association analysis of reproductive traits in the Finnish Landrace pig breed. Genetics Selection Evolution. 43:42-49.

Vallet, J.L., Freking, B.A., Leymaster, K.A., and Christenson, R.K. 2005. Allelic variation in the erythropoietin receptor gene is associated with uterine capacity and litter size in swine. Animal Genetics. 36:97-103.

Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G., Georges, M., and Andersson, L. 2003. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. Nature. 425:832-836.

Vincent, A.L., Short, T.H., Southwood, O.I., Plastow, G.S., Tuggle, C.K., and Rothschild, M.F. 1998. The prolactin receptor gene is associated with increased litter size in pigs. In: Proceedings of the Sixth World Congress on Genetics Applied to Livestock Production, Armidale, vol. 26, pp. 403-409.

Werzner, A., Pavlidis, P., Ometto, L., Stephan, W., and Laurent, S. 2013. Selective sweep in the *Flotillin-2* region of European *Drosophila melanogaster*. PLoS ONE. 8(2):e56629.

Wilson, E.R. and Johnson, R.K. Comparison of three-breed and backcross swine for litter productivity and postweaning performance. Journal of Animal Science. 52(1):18-25.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics. 15:323-354.

Zhang, H., Wang, S.Z., Wang, Z.P., Da, Y., Wang, N., Hu, X.X., Zhang, Y.D., Wang, Y.X., Leng, L., Tang, Z.Q., and Li, H. 2012. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC Genomics. 13:704-719.

Zimmerman, D.R. and Cunningham, P.J. 1975. Selection for ovulation rate in swine: population, procedures, and ovulation response. Journal of Animal Science. 40(1):61-69.

CHAPTER 2: Long-Term Selection for Litter Size in Swine Results in Shifts in Allelic Frequency in Regions Involved in Reproductive Processes

INTRODUCTION

Litter size is a composite trait influenced by many factors, including ovulation rate, uterine capacity, and embryonic survival. This trait is economically important and has been emphasized in selection programs in all modern maternal lines of swine. However, genetic progress can be difficult to achieve, as litter size is characterized by low heritability (Bidanel, 2011) and influenced by many loci, each with relatively small effects (Pomp *et al.*, 2001).

However, response in Nebraska Index Line (NIL), a composite population of pigs based on Landrace and Large White genetics that was subjected to long-term selection for litter size beginning in 1981, proved that genetic improvement is possible (Hsu, 2011). In the first 11 generations, NIL was selected for increased ovulation rate (OR) and embryonic survival (ES) combined in an index. Selection was performed for increased number of fully formed pigs per litter (FF) in the following three generations (12-14) and increased number of pigs born alive per litter (NBA) and birth weight (BW) in the next four generations (15-19) to correct for an increase in stillborn piglets. To increase lean growth in addition to increasing NBA, the objective of selection included increased growth rate (WT), decreased backfat (BF), and increased *Longissimus* muscle area (LMA) in all subsequent generations. Generation interval was one year. A control line (CTRL) from the same Landrace by Large White founder population in which no selection was practiced was maintained throughout the selection period. Full- and half-sibling matings were avoided in all generations in both lines.

All selection strategies, both direct and indirect, were successful in increasing NBA. In the most recent generation, the average NBA reached 13.4 piglets per litter in NIL, while ranging

between eight and nine piglets per litter throughout the experiment in CTRL. After 28 generations of selection, genetic variance in NBA was still present, indicating potential to continue to increase NBA through selection. A more detailed description of the lines, selection strategies, and results can be found in Hsu (2011).

In contrast to maternal lines, selection for litter size is not practiced in commercial paternal lines, including common breeds such as Duroc and Hampshire. On average, total number of piglets born per litter (TNB) is 9.9 in Duroc, compared to 14.2 and 14.6 in Large White and Landrace, respectively (Bidanel, 2011). Rather, traits influencing growth and meat quality are emphasized in paternal line selection programs.

The selection strategy utilized in NIL was unique and not practiced in commercial populations, particularly in the first 11 generations when selection was based on an index of OR and ES. With the unparalleled long-term selection that has been performed on NIL and the contrast of a control line of the same origin and a Duroc by Hampshire paternal cross (Petry *et al.*, 2005), the resource populations used in this study provide a unique opportunity to use genomic information to identify polymorphisms and genes that have been targeted by selection.

The objective of this study was to integrate four complementary selective sweep identification methods with litter size QTL detection to uncover regions of the genome that displayed shifts in allelic frequency in NIL compared to populations in which selection for litter size was not practiced. We hypothesized that regions where selection triggered changes in allelic frequency will be detected by multiple methods and overlap with QTL for litter size.

MATERIALS AND METHODS

Resource populations

The experimental populations include representative males and females sampled from NIL (n = 74, generation 32) and CTRL (n = 61, generation 30) and females from a commercial Duroc by Hampshire cross (DxH) population (n = 92). Genome-wide association studies (GWAS) and heritability estimates were conducted with females from the University of Nebraska – Lincoln (UNL) sow reproductive longevity resource population (hereafter "UNL resource population"), based on NIL and commercial Large White and Landrace genetics. Sows from this population (n = 481-701) were phenotyped for an array of developmental and reproductive phenotypes (Tart *et al.*, 2013).

DNA isolation and genotyping

DNA was isolated from tail clips using DNeasy blood and tissue kits (Qiagen). High-density genotypes were obtained using Porcine SNP60 BeadArray (Illumina). All SNPs with a GenCall genotype quality score of less than 0.40 and a sample and SNP genotyping call rate below 0.80 were removed to ensure quality. The remaining SNPs (n = 55,619) were mapped on build 10.2 of the reference assembly of the porcine genome.

Genetic diversity statistics

Variation and diversity within populations was estimated using the proportion of polymorphic SNPs, SNP heterozygosity, and the population inbreeding estimate, F_{IS}. GENEPOP software was used to calculate F_{IS} after pruning SNPs with minor allele frequency (MAF) < 0.01 and genotyping rate < 0.95. Locus-by-locus deviation from Hardy-Weinberg equilibrium (HWE) within each population was assessed using a chi-square test with the significance threshold corrected for the number of tests (alpha = 0.05 / number of SNPs = $9.8*10^{-7}$). Variation among

populations across the genome was estimated based on 1) allelic frequency differences between populations, and 2) changes in SNP heterozygosity. Variation in allelic frequencies between populations was estimated using chi-square contingency tests comparing CTRL and DxH to NIL (alpha = 0.0001 / number of SNPs = $2.0*10^{-9}$) and Wright's fixation index (F_{ST}) between CTRL and NIL. Unmapped SNPs and 1-Mb windows with three SNPs or less were excluded from these analyses. The ratio of SNP heterozygosity between NIL and CTRL, calculated in sliding windows of ten consecutive SNPs, was used to measure relative reduction in heterozygosity across the genome. Clustering among individuals was assessed by multidimensional scaling (MDS) in Plink (Purcell *et al.*, 2007) based on pairwise identity by state of 21,524 autosomal SNPs after removal of SNPs with across population minor allele frequency < 0.01, genotyping call rate < 0.95, and $r^2 > 0.2$.

Genome-wide association analyses

The proportion of phenotypic variance for litter size traits, NBA at parity one (P1) and parity two (P2) and TNB-P1 and -P2, explained by each 1-Mb window was estimated from high-density SNP genotypes from UNL resource population females using a Bayes B approach implemented via the GENSEL software package (Fernando and Garrick, 2008). The analyses were performed using 41,000 iterations, with a burnin of 1,000 iterations, meaning the first 1,000 iterations were discarded. The π value, or proportion of markers assumed to have no effect on the trait of interest, was set to 0.99, and line, batch, and diet were included in the model as fixed effects. Heritability was estimated using a sire model, including batch and diet as fixed effects and sire and litter as random effects. Additive genetic variance was calculated as four times the sire variance and divided by the total variance to obtain heritability estimates.

Simulation of Wright's fixation index

The F_{ST} statistic was first described by Wright (1951) and is a measure of genetic differentiation between populations or subpopulations. It is calculated as $F_{ST} = 1 - \frac{H_S}{H_T}$, where $H_S = \frac{(2pq_{CTRL}) + (2pq_{NIL})}{2}$ and $H_T = 2pq_{total}$. The F_{ST} analysis was restricted to 41,835 SNPs mapped to autosomes and the X chromosome that were polymorphic in either NIL or CTRL or monomorphic for a different allele in each population. Across this set of SNPs, the average MAF was 0.23, and the median observed F_{ST} was 0.048. Simulation of Wright's fixation index (F_{ST}) distribution under a neutral model was conducted to distinguish allelic frequency differences between NIL and CTRL most likely driven by selection from those that are a result of genetic drift. The simulation was conducted with a biallelic locus segregating at G_0 with a MAF of 0.23 in a population of 30 males and 95 females that was allowed to drift for 30 generations. This reflects the average number of individuals selected each generation in NIL and CTRL and the approximate number of generations they were subjected to selection. The process was repeated 41,835 times, allowing changes in allelic frequencies. The F_{ST} was calculated after each round of simulation, and the simulated F_{ST} distribution was compared with the observed F_{ST} values.

Relative Extended Haplotype Homozygosity (REHH)

The extended haplotype homozygosity (EHH) statistic is used to identify regions targeted by artificial selection where allelic frequencies have increased faster than would be expected via drift (Gärke *et al.*, 2014). SNP genotypes for all NIL individuals were phased into haplotypes using FastPHASE (Scheet and Stephens, 2006) after removal of SNPs with a MAF < 0.01 and a genotyping call rate of < 0.95. Ten random starts were used for each chromosome. Chromosomes were grouped by size to test for the optimal number of haplotype clusters (K). A K of 35 was used for all chromosomes. Haplotype cores were identified using Sweep software

(Sabeti *et al.*, 2002) based on genetic distance estimated from recombination rates (marker H) of 0.1. This value is higher than the default setting of 0.04 due to the increase in haplotype block length in pigs compared to humans for which the program was designed. Relative extended haplotype homozygosity (REHH) was calculated for each haplotype of each core as $EHH_t/\overline{EHH}$.

EHH_t is the EHH value of core haplotype t and is calculated as $EHH_t = \frac{\sum_{i=1}^{s} \binom{e_{ti}}{2}}{\binom{e_t}{2}}$, where c is the number of samples of a particular core haplotype, e is the number of samples of a particular extended haplotype, and s is the number of unique extended haplotypes. $\overline{EHH}$ is the decay of EHH on other core haplotypes combined and is calculated as $\overline{EHH} = \frac{\sum_{j=1, j \neq t}^{n} \left[\sum_{i=1}^{s} \binom{e_i}{2} \right]}{\sum_{i=1, i \neq t}^{n} \binom{e_i}{2}}$. The p-values corresponding to each core haplotype were calculated in 20 bins according to haplotype frequency (q = 0 – 0.05, q = 0.05 – 0.1, q = 0.1 – 0.15, etc.).

We analyzed core haplotypes in NIL consisting of at least nine SNPs with significant REHH values ($P < 0.05$) and a haplotype frequency greater than 0.28. Haplotype size was set to nine SNPs because shorter haplotypes are likely to be a result of drift rather than selection. Based on the same Porcine SNP60 BeadArray genotypes of six commercial lines, Veroneze *et al.* (2013) estimated the average linkage disequilibrium block size to be 400 kb in pigs. The average space between SNPs on the Porcine SNP60 BeadArray (Illumina) is 43.4 kb (http://res.illumina.com/documents/products/datasheets/datasheet_porcinesnp60.pdf); therefore, nine SNPs encompass around 400 kb. The mean haplotype frequency in NIL was 0.15 with a standard deviation of 0.13. We expected haplotypes under strong long-term selection to

have a frequency above average, so we set the haplotype frequency threshold at the mean plus one standard deviation (0.28).

Gene Ontology

Significant core haplotypes and windows that are responsible for significant variation between populations and explain the largest proportion of the phenotypic variance of the four litter size traits analyzed were extended by 0.5 Mb in both directions for functional characterization of positional candidate genes using the *Sus scrofa* build 10.2. The BIOMART tool in the Ensembl database (http://uswest.ensembl.org/biomart/martview/8aae4d40f586eab6d5336a79044e6bf1) was used to identify positional candidate genes and their gene ontology terms. Functional annotation, gene ontology term enrichment, and pathway analyses were performed on the human orthologs using DAVID (http://david.abcc.ncifcrf.gov).

Single-marker association

Single-marker association analyses for litter size and other reproductive traits were performed in JMP 11 with genotypes and phenotypes from UNL Resource Population females using additive and dominance general linear mixed models. The models included batch, diet, and genotype as fixed effects and litter and sire as random effects. Pairwise comparisons between genotype least squares means (LSM) were based on the Tukey test.

Linkage disequilibrium analysis

Linkage disequilibrium was assessed in a candidate region located on SSC2 (12-16 Mb) using Porcine SNP60 BeadArray genotypes from UNL Resource Population females and Haploview (Barrett *et al.*, 2005). This region contained two SNPs with F_{ST} within the 99.9

percentile of F_{ST} simulation data, four positional candidate genes, and a QTL for litter size. Genotypes from all SNPs in the region included on the Porcine SNP60 BeadArray were extracted from UNL Resource Population females. Linkage disequilibrium between each pair of SNPs was calculated as $r^2 = \frac{D^2}{p_A q_A p_B q_B}$, where $D = ru - st$. Variables r and u are the observed gametic frequencies of the coupling gametes, A_1B_1 and A_2B_2, respectively. Variables s and t are the observed gametic frequencies of the repulsion gametes, A_1B_2 and A_2B_1, respectively. A and B represent the two loci between which linkage disequilibrium is being calculated.

cDNA sequencing and SNP discovery

Four positional candidate genes located near two SNPs that are within the 99.9 percentile of F_{ST} simulation data and next to a QTL for litter size were sequenced for SNP discovery. Extraction of RNA from ovarian tissue of UNL resource population females, spleen of DxH females, and blood of NIL females was accomplished using TRIZol (Life Technologies), RNeasy kits (Qiagen), or Tempus Tube Spin RNA Isolation Kits (Life Technologies), followed by first-strand cDNA synthesis (GE Healthcare). Four total pools were made with cDNA of seven to nine non-littermates from NIL, DxH, and UNL resource population (n = 2), with each individual equally represented in the pool. Each gene was amplified in each pool using GoTaq Flexi DNA polymerase (Promega), and the PCR products were treated with ExoSAP-IT (USB Corporation). Sequencing was carried out using dye terminators on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems). Sequencer software (Gene Codes) was used to align the sequences and identify polymorphisms. Two informative polymorphisms were genotyped using KASPar (K Biosciences) in UNL Resource Population females with small (n = 75) and large (n = 75) TNB-P2 phenotypes adjusted with TNB-P2 conditional residuals. The model to obtain conditional residuals included batch and diet as fixed effects and litter and sire as random effects. The total

number of individuals successfully genotyped for the exon 4 and exon 7 SNPs were 111 and 120, respectively. Association analyses of these SNPs with TNB-P2 was carried out in JMP 11 using additive linear mixed models. Batch, diet, and genotype were fixed effects and litter and sire were included as random effects. Pairwise comparisons between genotype least square means (LSM) were based on the Tukey test. Additive genetic variance explained by SNPs was calculated using the regression coefficient from the single marker association analysis as α and allelic frequencies with the equation $2pq\alpha^2$ (Falconer and Mackay, 1996). Additive genetic variance was divided by total variance obtained from the single marker association analysis to get the percentage of TNB-P2 variance explained by each SNP.

RESULTS AND DISCUSSION

Genetic diversity

The DxH population had the greatest proportion of polymorphic SNPs (83.3%), followed by CTRL (76.2%); NIL had the lowest proportion of polymorphic SNPs (73.4%), likely due to a reduction in effective population size (N_e), genetic drift, or selection that led to an increase in the frequency of favorable alleles above what would be expected under a neutral model (Table 1). Interestingly, initial validation studies of the Porcine SNP60 BeadArray found that Duroc and Hampshire had fewer polymorphic loci (n = 41,846 and 43,496, respectively) included in the chip than Large White and Landrace (n = 51,447 and 49,946, respectively); in addition, the mean MAF was greater in Large White (0.26) and Landrace (0.24) than in Duroc (0.20) and Hampshire (0.20) (http://res.illumina.com/documents/products/datasheets/datasheet_porcinesnp60.pdf). However, in our study, the crossbred status of the DxH population is reflected in the larger expected heterozygosity and proportion of polymorphic SNPs compared to NIL and CTRL.

All three populations had negative F_{IS} values; therefore, none of the populations sampled show evidence of inbreeding. Distinct clustering of the three populations illustrates the results of selective pressure in NIL compared to CTRL and population differences between the Landrace by Large White composites and the Duroc by Hampshire cross (Figure 1). NIL had the most SNPs that deviated significantly from HWE (n = 47), and CTRL had the least (n = 33) with DxH intermediate (n = 42). While most of the SNPs that deviated significantly from HWE were not located near each other, one region on SSC16 (76-77 Mb) contained six SNPs that were not in HWE in NIL, yet were in HWE in the other populations. This region includes the gene *GLRA1* (SSC16, 77.3 Mb), known for involvement in fertilization (geneontology.org).

Genome-wide association analyses

Litter size traits are characterized by low heritability and complex genetic determinism (Bidanel, 2011). In the UNL resource population, heritability was 0.16 for NBA-P1 and 0.01 for NBA-P2. Heritability of TNB was estimated at 0.12 and 0.08 for -P1 and -P2, respectively. Genome-wide association studies determined that combined SNP effects explained 7.6%, 2.5%, 3.7%, and 3.1% of the phenotypic variation for NBA-P1, NBA-P2, TNB-P1, and TNB-P2, respectively (Table 2; Figure 2). The three 1-Mb windows that explained the largest proportion of genetic variation for each litter size trait are shown in Table 3. A major window located on SSC13 (36-37 Mb) explained 0.26% of the variation for NBA-P1 and 0.10% of the variation for TNB-P1. Other major windows that explained genetic variation of multiple litter size traits explained a small proportion of the variation (0.10 - 0.14 %) of two or three traits (Table 4). Regions on SSC7 (15-16 Mb) and SSC9 (4-5 Mb) explained a large proportion of variation for NBA-P1, 0.30% and 0.95%, respectively, but less than 0.09% of the variation for the other three litter size traits.

Relative Extended Haplotype Homozygosity

Twenty-six core haplotypes had significant REHH p-values ($P < 0.05$), contained at least nine SNPs, and had a haplotype frequency at or above 0.28. These haplotypes are presented in Table 5, along with candidate genes located within them and their functions (geneontology.org). We expect SNPs with large effects on litter size to be under stronger selection and, therefore, increase in frequency faster than SNPs with smaller effects. The surrounding haplotype block will most likely be longer, as recombination will not have had time to break linkage with nearby SNPs. Of the 26 significant core haplotypes above the frequency threshold, the longest haplotype was located on SSC9 (15.9-17.5 Mb) and was 43 SNPs in length. The average haplotype length was 16.6 SNPs. The haplotype on SSC9 (53.7-55.0 Mb) had the greatest

frequency, at 0.56. The core haplotype on SSC12 (58.4-59.4 Mb) is located near a SNP with an F_{ST} within the 99.9 percentile of F_{ST} simulation data (SSC12, 59.9 Mb) and a litter size QTL (SSC12, 60-61 Mb). Three other core haplotypes are located next to windows identified via contingency tests as having a high proportion of SNPs with significantly different allelic frequencies between NIL and CTRL or NIL and DxH. These cores are located on SSC6 (34.4-35.4 Mb and 36.2-36.7 Mb) and SSC12 (10.7-11.1 Mb).

Significant shifts in genetic structure between populations

Shifts in genetic structure across SNPs and genomic windows were detected by analysis of loss in heterozygosity and changes in allelic frequencies between NIL and the other populations. Heterozygosity levels in sliding windows of ten consecutive SNPs were compared between NIL and CTRL. Eight chromosomal regions displayed windows with a greater than four-fold reduction in heterozygosity in NIL compared to CTRL. The most extended regions were located on SSC1 (95.9-97.0, 129.1-130.4 Mb) and SSC8 (59.0-60.9 Mb). Candidate genes in these regions include *IGFBP7* (SSC8, 59.0 Mb), known to play a role in embryo implantation and other processes related to pregnancy (geneontology.org).

Significantly different allelic frequencies between populations were present in 20.5% and 45.2% of SNPs in contingency test comparisons of NIL vs. CTRL and NIL vs. DxH, respectively ($P < 2.0*10^{-9}$). Some of the 1-Mb windows that were rich in SNPs with significantly different frequencies between populations overlapped with QTLs for litter size detected in the UNL resource population. Candidate genes with potential to influence fertility related traits were identified in these regions (Table 6; geneontology.org).

We acknowledge that an important proportion of the allelic frequency differences between NIL and CTRL are neutral and simply a result of genetic drift. To distinguish differences

in allelic frequency between NIL and CTRL across the genome that are most likely the result of selection rather than drift, we conducted a simulation of Wright's fixation index (F_{ST}) to generate an expected distribution under a neutral model. The simulation resulted in a median F_{ST} of 0.04, which is similar to the observed median F_{ST} of 0.048. There were 64 SNPs (0.15%) with F_{ST} greater than 0.616, which corresponds to the 99.9 percentile of the simulated data (Figure 3), with four SNPs (SSC2, 82.0 Mb; SSC8, 19.9 Mb; SSC12, 59.9 Mb; SSC17, 47.2 Mb) having observed F_{ST} values greater than the maximum simulated F_{ST} (0.812). Several candidate genes were located near these SNPs, including *NSD1* (SSC2, 82.1 Mb), *FGFR4* (SSC2, 82.4 Mb), *SLC34A2* (SSC8, 19.8 Mb), and *RBPJ* (SSC8, 20.4 Mb). These genes are all involved in some aspect of embryonic development. In addition, *NSD1* is involved in estrogen receptor binding and androgen receptor transactivation, while *RBPJ* is also involved in neuron differentiation (geneontology.org). Regions that include multiple SNPs with F_{ST} within the 99.9 percentile of the simulated F_{ST} data, overlap with litter size QTL or significant haplotypes, and/or were previously identified as regions with high differentiation between populations via loss in heterozygosity in NIL compared to CTRL or the NIL vs. CTRL contingency test are presented in Table 7 along with candidate genes with functions relating to reproductive processes identified within these regions (geneontology.org). The region on SSC2 (13-14 Mb) is particularly noteworthy as it is the only region containing multiple SNPs within the 99.9 percentile of F_{ST} simulation data located within close proximity to a major litter size QTL. Candidate genes identified in this region included *P2X3R* (SSC2, 13.3 Mb) and *SSRP1* (SSC2, 13.4 Mb; Table 6). The Porcine SNP60 BeadArray does not contain any SNPs in the adjacent 1-Mb window directly following the SNPs with F_{ST} within the 99.9 percentile of F_{ST} simulation data (SSC2, 14-15 Mb). When SNPs resume around 15.6 Mb, F_{ST} values remain high, though not within the 99.9 percentile. Therefore, the region was extended from 13-14 Mb to 13-16 Mb for candidate gene identification. Two

additional candidate genes were identified: *PTPRJ* (SSC2, 15.6 Mb) and *NUP160* (SSC2, 15.9 Mb). Both of these genes are involved in immune response, and *PTPRJ* also impacts embryogenesis (geneontology.org).

Another region on SSC2 (43.8-44.3 Mb), containing a cluster of four SNPs with F_{ST} within the 99.9 percentile of simulation data, was located near a litter size QTL identified by Lei *et al*. (2011). A mutation on SSC2 (44.3-44.5 Mb) in miR-27a gene was significantly associated with TNB and NBA across parities in a Chinese line of pigs (Lei *et al.*, 2011). However, the QTL did not have a significant effect in Large White pigs in the study by Lei *et al.* (2011), nor was it identified as a major QTL for litter size in our study.

Single marker association

The 1-Mb window on SSC2 (12-13 Mb) near two SNPs with F_{ST} within the 99.9 percentile of F_{ST} simulation data explains the most genetic variance of TNB-P2. Furthermore, this window ranks 6th in NBA-P2 genetic variance explained and 9th in TNB-P1 variance explained. Of all SNPs in the window, *ASGA0095946* (SSC2, 12.46 Mb) explained the most variance of TNB and NBA -P2, and *ALGA0118548* (SSC2, 12.51 Mb) explained the most variance of TNB-P1. Single-marker association analyses were performed for TNB and NBA -P1 and -P2 for both of these SNPs as well as *ALGA0111847* (SSC2, 13.38 Mb) and *ASGA0097301* (SSC2, 13.45 Mb), the two SNPs with F_{ST} within the 99.9 percentile of F_{ST} simulation data. P-values for additive and dominance models as well as least squares means and standard errors for each genotype can be found in Table 8. *ASGA0097301* was significantly associated with TNB-P1 ($P < 0.05$) and suggestively associated with TNB-P2 ($P < 0.2$) under an additive model and suggestively associated with both traits under a dominance model ($P < 0.15$). *ASGA0095946* was suggestively associated with NBA-P2 (P

> 0.1) and TNB-P2 (P = 0.1) under an additive model. There were no significant differences between genotype least squares means of any SNP for any trait.

Linkage disequilibrium analysis

Linkage disequilibrium was analyzed in the candidate region on SSC2 (12-14 Mb). The largest LD block was nearly 200 kb long and was located on SSC2 12.7-12.9 Mb. Other LD blocks were much smaller in size. Linkage disequilibrium may exist between the top SNP in the QTL region for TNB and NBA -P2 (*ASGA0095946*; SSC2, 12.46 Mb) and the top SNP in the QTL region for TNB-P1 (*ALGA0118548*; SSC2, 12.51 Mb) as r^2 between the two SNPs was 0.79; however, there are SNPs between them with low LD. The two SNPs with F_{ST} within the 99.9 percentile of F_{ST} simulation data (*ALGA0111847*; SSC2, 13.38 Mb and *ASGA0097301*; SSC2, 13.45 Mb) exhibited low linkage, as r^2 between them was 0.23. There is no evidence that the QTL from 12-13 Mb is linked to the F_{ST} SNPs as r^2 between the top two QTL SNPs and two F_{ST} SNPs range between 0 and 0.14. However, the SNPs used were sparse; with more SNPs, the QTL may have shifted toward the SNPs with high F_{ST} and candidate genes located near them. These SNPs show huge differences in allelic frequency between NIL and CTRL, and promising candidate genes are located within them. Therefore, we proceeded to study this region in further detail.

Positional candidate gene analysis

Four positional candidate genes located in the region described above that included multiple SNPs with F_{ST} within the 99.9 percentile of F_{ST} simulation data and a QTL for litter size were selected for further analysis. P2X purinoceptor (*P2X3R*; SSC2, 13.3 Mb), structure specific recognition protein 1 (*SSRP1*; SSC2, 13.4 Mb), and nucleoporin 160kDa (*NUP160*; SSC2, 15.9 Mb) were fully sequenced and protein tyrosine phosphatase receptor type J (*PTPRJ*; SSC2, 15.6 Mb) was partially sequenced for polymorphism discovery. Six, two, and four synonymous SNPs were

identified in *P2X3R*, *SSRP1*, and *PTPRJ*, respectively. No SNPs were identified in *NUP160*. All SNPs identified in *SSRP1* and *PTPRJ* were fixed in NIL, polymorphic in the UNL resource population, and fixed or nearly fixed for the same allele in DxH. However, the most interesting difference in allelic frequency across populations was identified in *P2X3R*; two SNPs were fixed in NIL and polymorphic in DxH and UNL resource population with nearly equal frequencies.

The P2X3R protein is a member of a family of membrane ion channels that open in response to the binding of extracellular ATP. They operate in multimers; P2X3R multimerizes with itself or P2X2R (North, 2002). Both P2X3R homomers and P2X2R/P2X3R heteromers are associated with response to pain stimuli (Cockayne *et al.*, 2005). The family of P2X receptors have been detected in the uterine epithelium of rats, and their expression greatly increases just prior to implantation, suggesting a role in the preparation of the uterine epithelium for implantation and pregnancy (Slater *et al.*, 2000, 2002a, 2002b). In addition, P2X3R is involved in the CREB pathway, which is associated with social learning and memory formation (Benito and Barco, 2010; Sakamoto *et al.*, 2011; Chen *et al.*, 2012). While P2X2R mediates the initial response, the P2X2R/P2X3R heteromer, along with P2X7R, is responsible for maintaining oxytocin and vasopressin release in response to ATP and phenylephrine stimulus in the hypothalamus (Gomes *et al.*, 2009). Oxytocin fulfills various roles in parturition, milk letdown, and maternal and copulatory behavior. In addition to regulating water balance and blood pressure, vasopressin is involved in social and sexual behavior (Hadley and Levine, 2007). Vasopressin's V1a receptor, AVPR1A, mediates vasopressin's social and sexual behavior functions (geneontolgy.org; Hadley and Levine, 2007; Walum *et al.*, 2008; Hammock and Young, 2005). Polymorphisms in *AVPR1A* were associated with both age at puberty (AP) and lifetime number of parities (LTNP) in pigs (Tart *et al*., 2013; Lucot *et al.*, 2015). In addition, *P2X3R* is located in a previously identified QTL region for AP (Trenhaile, unpublished data). AP is

negatively correlated with LTNP, lifetime NBA, and lifetime TNB (Tart *et al.*, 2013). Negative correlations also exist between genomic prediction values of AP and these lifetime reproductive traits (Lucot *et al.*, 2015). Single-marker association analysis for *ALGA0111847*, the closest SNP to *P2X3R* on the SNP60 BeadArray and one of two SNPs in this region with an F_{ST} in the 99.9 percentile of F_{ST} simulation data, indicated a significant association between this marker and AP under both an additive and a dominance model ($P < 0.01$). The *GG* genotype of this SNP is significantly associated with earlier age at puberty than the *GT* genotype ($P < 0.01$). Individuals with the *GG* genotype also tend to have greater TNB and NBA in both parities 1 and 2 than individuals with *GT* and *TT* genotypes, though these effects were not significant (Table 8).

The two synonymous SNPs identified by sequencing that were fixed in NIL and polymorphic in DxH were located in exons 4 and 7 of the *P2X3R* gene. These SNPs were in high linkage disequilibrium ($r^2 = 0.79$). Due to the influence of NIL in the UNL Resource Population, only five and four individuals were identified that were homozygous for the alternate (non-NIL) allele in exon 4 and exon 7, respectively. Therefore, association analyses were run including and excluding these individuals due to the large standard error associated with the least squares means of the rare genotype. Neither SNP had a significant association with TNB-P2 (Table 9), though the exon 7 SNP had a suggestive association when individuals with the *TT* genotype were included ($P < 0.07$). No significant differences existed between least squares means of any two genotypes. However, a trend is visible for both SNPs in which the *GG* and *CC* genotypes, which are fixed in NIL, are favorable and the *AA* and *TT* genotypes are unfavorable (Figures 4 and 5). As these are synonymous SNPs, they are not true causative mutations that affect litter size. However, this does not completely discount their value as they may be in linkage disequilibrium with a regulatory mutation that affects gene expression and influences litter size.

CONCLUSION

Genetic structure differences in NIL compared to populations in which selection for litter size was not practiced were detected via loss in heterozygosity in NIL and changes in allelic frequencies between populations across the swine genome. Haplotypes likely to be under selection in NIL were also identified. GWAS uncovered common QTL regions that explained phenotypic variation across litter size traits. Several regions displaying significant shifts in allelic frequencies overlap with major QTL. Candidate genes located in these regions involved in major reproductive functions were identified, though much of the phenotypic variance resulting in increased litter size in NIL could be a result of variation in regulatory regions in addition to differences in coding regions focused on in this work. It is likely that these differences in allelic frequency may be due to selection for increased litter size in NIL. Many of the selective sweep regions and candidate genes may be valuable for additional validation and consideration in future selection programs for increased litter size and reproductive traits.

LITERATURE CITED

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps**.** Bioinformatics*.* 21(2):263-265.

Benito, E. and Barco, A. 2010. CREB's control of intrinsic and synaptic plasticity: implications for CREB-dependent memory models. Trends in Neurosciences. 33(5):230-240.

Bidanel, J.P. 2011. Biology and genetics of reproduction. In *The Genetics of the Pig.* 2nd Edition. Edited by Rothschild, M.F. and Ruvinsky, A. Wallingford, UK: CAB International; 218-241.

Chen, D.Y., Bambah-Mukku, D., Pollonini, G., and Alberini, C.M. 2012. Glucocorticoid receptors recruit the CaMKIIα-BDNF-CREB pathways to mediate memory consolidation. Nature Neuroscience. 15(12):1707-1714.

Cockayne, D.A., Dunn, P.M., Zhong, Y., Rong, W., Hamilton, S.G., Knight, G.E., Ruan, H.Z., Ma, B., Yip, P., Nunn, P., McMahon, S.B., Burnstock, G., and Ford, A.D.P.W. 2005. $P2X_2$ knockout mice and $P2X_2/P2X_3$ double knockout mice reveal a role for the $P2X_2$ receptor subunit in mediating multiple sensory effects of ATP. J. Physiol. 567(2):621-639.

DAVID Functional Annotation Bioinformatics Microarray Analysis http://david.abcc.ncifcrf.gov.

Ensembl Genome Browser BioMart http://uswest.ensembl.org/biomart/martview/8aae4d40f586eab6d5336a79044e6bf1.

Falconer, D.S. and Mackay, T.F.C. 1996. Introduction to Quantitative Genetics. Fourth edition. Longman Group Ltd.

Fernando, R.L. and Garrick, D.J. 2008. Gensel User Manual for Portfolio of Genomic Selection Related Analyses. Animal Breeding and Genetics. Iowa State University, Ames.

Gärke, C., Ytournel, F., Sharifi, A.R., Pimentel, E.C.G., Ludwig, A., and Simianer, H. 2014. Footprints of recent selection and variability in breed composition in the Göttingen Minipig genome. Stichting International Foundation for Animal Genetics. 45:381-391.

Gene Ontology Consortium http://geneontology.org.

Gomes, D.A., Song, Z., Stevens, W., and Sladek, C.D. 2009. Sustained stimulation of vasopressin and oxytocin release by ATP and phenylephrine requires recruitment of desensitization-resistant P2X purinergic receptors. Am. J. Physiol. Regul. Integr. Comp. Physiol. 297:R940-R949.

Hadley, M.E. and Levine, J.E. 2007. Endocrinology. 6th Edition. Pearson.

Hammock, E.A.D. and Young, L.J. 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. Science. 308:1630-1634.

Hsu, W.L. 2011. Analysis of long-term selection (28 generations) for reproduction, growth, and carcass traits in swine. PhD dissertation. University of Nebraska – Lincoln, Animal Science Department.

Lei, B., Gao, S., Luo, L.F., Xia, X.Y., Jiang, S.W., Deng, C.Y., Xiong, Y.Z., and Li, F.E. 2011. A SNP in the miR-27a gene is associated with litter size in pigs. Mol. Biol. Rep. 38:3725-3729.

Lucot, K.L., Spangler, M.L., Trenhaile, M.D., Kachman, S.D., and Ciobanu D.C. Evaluation of reduced subsets of single nucleotide polymorphisms for the prediction of age at puberty in sows. Animal Genetics. 46(4):403-409.

North, R.A. 2002. Molecular physiology of P2X receptors. Physiol. Rev. 82:1013-1067.

Petry, D.B., Holl, J.W., Weber, J.S., Doster, A.R., Osorio, F.A., and Johnson, R.K. 2005. Biological responses to porcine respiratory and reproductive syndrome virus in pigs of two genetic populations. Journal of Animal Science. 83(7):1494-1502.

Pomp, D., Caetano, A.R., Bertani, G.R., Gladney, C.D., and Johnson, R.K. 2001. Applying functional genomics research to the study of pig reproduction. Reproduction. Suppl. 58:277-292.

PorcineSNP60 v2 Genotyping BeadChip http://res.illumina.com/documents/products/datasheets/datasheet_porcinesnp60.pdf.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-75.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., and Lander, E.S. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. 419:832-837.

Sakamoto, K., Karelina, K., and Obrietan, K. 2011. CREB: a multifaceted regulator of neuronal plasticity and protection. Journal of Neurochemistry. 116:1-9.

Scheet, P. and Stephens, M. 2006. A fast and flexible statistical model for large-scale population data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78(4):629-644.

Slater, N.M., Barden, J.A., Murphy, C.R. 2000. Distributional changes of purinergic receptor subtypes (P2X 1-7) in uterine epithelial cells during early pregnancy. Histochem. J. 32(6):365-372.

Slater, N.M., Murphy, C.R., and Barden, J.A. 2002a. Tenascin, E-cadherin, and P2X calcium channel receptor expression is increased during rat blastocyst implantation. Histochem. J. 34(1-2):13-19.

Slater, N.M., Murphy, C.R., and Barden, J.A. 2002b. Purinergic receptor expression in the apical plasma membrane of rat uterine epithelial cells during implantation. Cell Calcium. 31(5):201-207.

Stavréus-Evers, A., Masironi, B., Landgren, B.M., Holmgren, A., Eriksson, H., and Sahlin, L. 2002. Immunohistochemical localization of glutaredoxin and thioredoxin in human endometrium: a possible association with pinopodes. Molecular Human Reproduction. 8(6):546-551.

Tart, J.K., Johnson, R.K., Bundy, J.W., Ferdinand, N.N., McKnite, A.M., Wood, J.R., Miller, P.S., Rothschild, M.F., Spangler, M.L., Garrick, D.J., Kachman. S.D., and Ciobanu, D.C. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Animal Genetics. 44:387-397.

Tsilchorozidou, T., Honour, J.W., and Conway, G.S. 2003. Altered cortisol metabolism in polycystic ovary syndrome: insulin enhances 5alpha-reduction but not the elevated adrenal steroid production rates. Journal of Clinical Endocrinology & Metabolism. 88(12):5907-5913.

Veroneze, R., Lopes, P.S., Guimarães, S.E., Silva, F.F., Lopes, M.S., Harlizius, B., and Knol EF. 2013. Linkage disequilibrium and haplotype block structure in six commercial pig lines. Journal of Animal Science. 91(8):3493-3501.

Walum, H., Westberg, L., Henningsson, S., Neiderhiser, J.M., Reiss, D., Igl, W., Ganiban, J.M., Spotts, E.L., Pedersen, N.L., Eriksson, E., and Lichtenstein, P. 2008. Genetic variation in the vasopressin receptor 1a gene (AVPR1A) associates with pair-bonding behavior in humans. PNAS. 105(37):14153-14156.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics. 15:323-354.

CHAPTER 3: *AVPR1A* Genotypes and Genotype by Diet Interactions Influence Age at Puberty and Reproductive Longevity

INTRODUCTION

Sow reproductive longevity is an economically important trait as a longer reproductive life results in greater returns on the replacement gilt investment. However, many sows fail to reach the three parities required to cover development and maintenance costs of breeding females (Stalder *et al.*, 2003). Furthermore, reproductive failure accounts for a substantial proportion of involuntary culling decisions and is the most frequent reason for early culling (Mote *et al.*, 2009). In addition to an economic problem, sows unable to handle the rigors of modern swine production are an animal welfare concern. Therefore, the industry would greatly benefit from sows who were able to consistently provide reproductive success until voluntary culling.

Reproductive longevity is a composite trait, influenced by many factors including age at puberty, ovulation rate, age at first service, conception rate, number of piglets born alive and weaned, wean-to-service interval, and weight loss during lactation (Tart *et al.*, 2013). It is a polygenic trait, influenced by many genes with small effects, and is also largely determined by environmental factors. Phenotypic selection for reproductive longevity can be very difficult due to low heritability, sex-limited expression, and expression late in life. Age at puberty (AP) is negatively correlated with reproductive longevity and can be used as an early indicator of future reproductive performance (Stalder, 2004; Serenius and Stalder, 2006; Patterson *et al.*, 2010; Knauer *et al.*, 2011; Tart *et al.*, 2013). Both of these traits are dependent on the function of the hypothalamic-pituitary-gonadal axis, and their variation is expected to be influenced by the

same genes. Age at puberty is a moderately heritable trait, but traditional selection is still difficult as it requires the tedious, labor-intensive process of daily estrus detection via boar exposure. Marker-assisted selection for AP and reproductive longevity may provide a better alternative, as markers can be used to make selection decisions early in life on both sexes, improving accuracy and increasing genetic progress.

A resource population was developed at the University of Nebraska – Lincoln to study genetic factors influencing sow reproductive longevity. Genome-wide association studies (GWAS) uncovered a region on SSC5 that is a pleiotropic source of variation of both AP and reproductive longevity. The main candidate gene in this region is arginine-vasopressin receptor 1A (*AVPR1A*; Tart *et al.*, 2013), known to influence sexual and social behavior in multiple species, including voles and humans (geneontolgy.org; Hammock and Young, 2005; Hadley and Levine, 2007; Walum *et al.*, 2008). A non-synonymous SNP in *AVPR1A*, *BGIS0007637*, was associated with both AP and reproductive longevity. The *AA* genotype of this SNP was associated with 5.8 day later expression of first estrus compared to the *GG* genotype ($P < 0.05$) and 3.6 day later expression than the *AG* genotype ($P < 0.09$). The *GG* genotype was associated with 0.53 and 0.33 more lifetime parities than the *AA* ($P < 0.01$) and *AG* ($P < 0.08$) genotypes, respectively (Tart *et al.*, 2013). An objective of this study was to further characterize *AVPR1A* to assess its efficiency as a selection marker for early AP and increased reproductive longevity.

Gilts are usually developed on an ad libitum feeding regimen in order to obtain adequate body weight and backfat at breeding (Miller *et al.*, 2011). However, calorie-restricted feeding has been shown to slow aging and increase longevity in mammalian and invertebrate species (Merry, 2002). Furthermore, Klindt *et al.* (1999, 2001a) observed no adverse effects of moderate energy restriction during gilt development on reproductive performance through first farrowing. Restricted feeding during gilt development could provide a way for producers to

reduce feed costs without compromising subsequent reproductive performance. However, these results may not hold true in all populations, as different genetic backgrounds are expected to vary in sensitivity to fluctuations in energy balance and, therefore, exhibit different nutrient-dependent responses with regards to AP and reproductive longevity. Our working hypothesis is that physiological response in AP and reproductive longevity is dependent on the interaction of genetic background and energy input. Thus, additional objectives were to characterize the effects of moderate energy restriction during gilt development on AP and reproductive longevity and utilize genome-wide association analysis to identify diet-dependent effects of DNA markers on these traits.

MATERIALS AND METHODS

University of Nebraska – Lincoln Sow Reproductive Longevity Resource Population

All procedures involving animals were approved by the University of Nebraska Institutional Animal Care and Use Committee. Gilts (n > 1,500) originated from commercial Large White by Landrace rotational cross or Nebraska Index Line (NIL) dams and commercial maternal Landrace sires. Nebraska Index Line has undergone long-term selection for litter size since 1981 (Hsu, 2011). The gilts were developed in 13 batches of around 120 animals each. Approximately eight sires were used to generate about 36 litters each batch from which project gilts were selected. Sires and dams were as evenly represented as possible in the selected gilts. Around 56 days of age, experimental gilts were assigned to pens based on birth date, sire, and litter in two climate-controlled rooms, where they were housed through the developmental period. Gilts received the same dietary treatment and management prior to the developmental period, beginning at approximately 120 days of age.

Developmental Period

The developmental period began when the average age of the gilts was around 120 days and continued until they were moved to the breeding barn at approximately 240 days of age. During this time, gilts were allocated to either a standard, ad libitum corn-soybean meal diet or an energy restricted diet. In batches 1-4, energy restriction was accomplished by hand-feeding half of the gilts 75% of the amount of feed consumed by the other half of the gilts which were fed ad libitum. In batches 5-6, the same corn-soybean meal ad libitum diet was fed along with two diets containing 20% dried distillers grains and solubles (DDGS). One DDGS diet was fed ad libitum, while the other was restricted to 80% of that which was consumed by the gilts fed ad libitum. Phase feeding was introduced in batch 7; gilts were fed either a standard corn-soybean

meal diet or a reduced energy corn-soybean meal diet containing 40% soy hulls in three phases. Each phase lasted for about six weeks. Both diets were fed ad libitum, but the gilts on the reduced energy diet still only consumed about 80% of the energy consumed by gilts on the standard diet. Phase feeding with the same two diets continued in batches 8-13, with the addition of a third diet beginning in batch 10 and continuing through batch 13. This diet had decreased levels of lysine while also containing 40% soy hulls and providing only 80% of the energy consumed by gilts on the standard diet. Across all batches, all restricted diets were fortified with nutrients to ensure that all gilts received proper levels of all nutrients required for developing gilts (NRC, 1998), with the exception of the lysine restricted diet added in batches 10-13.

Each pen of gilts was moved daily to a pen in an adjacent room and exposed to a boar for 15 minutes, beginning when the oldest gilt in the pen reached 130 days of age and continuing until all gilts in the pen expressed estrus twice or the average age of the gilts reached 240 days. The dates of first and second estrus were recorded for each gilt, and AP was defined as age in days on the first day observed in estrus. Every two weeks throughout the developmental period, gilts were weighed and probed with an Aloka 500V real-time ultrasound instrument equipped with a 3.5-MHz, 17-cm linear transducer (Corometrics Medical System, Inc.) in order to get back fat thickness (BF) and *longissimus* muscle area (LMA) measurements at the 10^{th} rib.

Tissue Collection

In batches 11-13, gilts were sacrificed just prior to the start of the developmental period around 120 days of age (n = 12 in total) and about 1/3 of the way through the developmental period around 165 days of age (n = 27 in total). Gilts were transported from the farm to the University of Nebraska Meat Lab the morning of slaughter, which began at 7:00 a.m.

Hypothalamus, anterior and posterior pituitary, ovarian cortex, granulosa cells, corpora lutea, and liver tissues were collected from each gilt and flash-frozen in liquid nitrogen. *Longissimus dorsi* was also collected 45 minutes, day 2, and day 9 post slaughter.

Breeding and Sow Management

Gilts were moved to the breeding barn around 240 days of age, and received the same diet from this point on. Due to space limitations, only approximately 100 gilts were bred each batch. Gilts remained in production through four parities, and were only culled for structural or health problems or reproductive failure. Reproductive failure occurred when estrus was not expressed during the three to four week breeding period or a litter was not conceived or farrowed after one breeding. Date and reason for culling was recorded. After breeding, sows were housed in individual gestation stalls until moving to farrowing stalls around day 109 of gestation. Sows were weighed and probed at the 10^{th} rib for BF using the same ultrasound equipment before moving to farrowing and after weaning. Total number born (TNB), number born alive (NBA), number stillborn (SB), and number of mummies (MUM) were recorded for each parity. Birth and weaning weights of all piglets were recorded. Some cross-fostering was performed in order to create more even litter sizes and increase piglet survival. Piglets fostered on and off as well as number weaned were also recorded for each litter. Sows nursed their litters for approximately 21 days, and feed intake during the lactation period was recorded.

DNA Isolation and Genotyping

Tail samples or ear notches were collected from all gilts at processing shortly after birth for DNA isolation using DNeasy or Puregene tissue kits (Qiagen). Quality and quantity of DNA was assessed using gel electrophoresis and a NanoDrop Spectrophotometer (Thermo Scientific). Each gilt was genotyped for 62,183 SNPs with the Porcine SNP60 BeadArray (Illumina). All genotypes with a quality score below 0.2 were removed and replaced with allelic frequencies.

Individual SNPs and samples with a call rate below 0.8 were removed, leaving 52,736 SNPs and 1,236 individuals for analyses.

Sequence Analysis

Conditional residual AP was calculated in JMP 11 using a linear mixed model, including batch and diet as fixed effects and litter and sire as random effects. The *AVPR1A* gene was completely sequenced in gilts exhibiting low (n = 8) and high (n = 8) conditional residual AP as well as Meishan females (n=8), a breed known for early expression of first estrus and large litter size (Bidanel, 2011). The Meishan samples were kindly provided by Dr. Gary Rohrer of the United States Meat Animal Research Center. Amplification was accomplished using Amplitaq Gold DNA polymerase (Applied Biosystems), and the PCR products were treated with ExoSAP-IT (USB Corporation). Sequencing was carried out using dye terminators on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems). Sequences were aligned and polymorphisms identified in Sequencer software (Gene Codes). Non-synonymous polymorphisms identified were genotyped in individuals with low (n = 150) and high (n = 150) conditional residual AP using KASPar (K Biosciences) or the Porcine SNP60 BeadArray (Illumina).

Gene Expression Analysis

RNA was extracted from anterior and posterior pituitary and ovarian tissue using TRIZol (Life Technologies) or RNeasy kits (Qiagen), followed by first strand cDNA synthesis (GE Healthcare). *AVPR1A* gene expression was analyzed with reverse-transcription quantitative PCR using TaqMan assays and universal PCR mastermix (Applied Biosystems). The analysis was done in triplicate and the three values averaged to get mean cycle threshold (Ct) values for each individual for both *AVPR1A* and a reference gene (*RPL32*). Any Ct values that were not within 0.5 of the other two replicate values were removed from the analysis. Individual samples without two Ct values within 0.5 of each other were not used in the analysis. The standard curve was

used to calculate amplification efficiency for both *AVPR1A* and the reference gene with the formula $Efficiency = -1+10^{-1/slope}$. Mean normalized expression (MNE) was calculated from efficiency and Ct values as $MNE = \frac{[(E_{reference})^{\overline{Ct_{ref}}}]}{[(E_{AVPR1A})^{\overline{Ct_{AVPR1A}}}]}$, where $E = 2^{efficiency}$ (Simon, 2003). The $\log_{10}$(MNE+1) was calculated. T-tests were performed to determine if expression differences existed between tissues, *AVPR1A* genotypes, pre- and post-pubertal individuals, and individuals expressing first estrus early or late at an alpha level of 0.05.

Genome-Wide Association Analysis for Interaction between Diet and Genotype

Genomic regions that interacted with diet to influence AP and LTNP were assessed via GWAS using a Bayes B approach implemented via the GENSEL software package (Fernando and Garrick, 2008). The SNP set was doubled, and one set was coded to identify main effects and the other coded to identify interaction effects between diet and genotype (Table 10). In the main effects SNP set, the 11 genotype was coded as -10, the 12 genotype was coded as 0, and the 22 genotype was coded as 10 in all individuals. In the interaction effects SNP set, individuals fed a standard, ad libitum diet were coded the same as in the main effects SNP set, but in individuals fed an energy restricted diet, the 11 genotype was coded as 10, the 12 genotype was coded as 0, and the 22 genotype was coded as -10. The analysis was performed using 41,000 iterations, with a burnin of 1,000 iterations. The π value was set to 0.99 for main effects and 0.995 for interaction effects. Line, batch, and diet were included in the model as fixed effects.

Statistics

The effect of AP and energy intake during development on the probability to generate parities 1 to 3 was assessed in R using a generalized linear mixed model, including AP as a covariate, batch and diet as fixed effects, and litter as a random effect. A generalized linear mixed model was also used to assess the direct effect of *AVPR1A G31E* genotype and the

interaction between this SNP and energy intake during development on the probability that a sow would produce up to three litters. Batch, diet, SNP, and SNP by diet were fixed effects, and litter was a random effect in this model.

Single-marker association analyses for *AVPR1A* SNPs and AP and LTNP were performed in JMP 11 using additive general linear mixed models and individuals that expressed first estrus early and late (n = 300). The models included batch, diet, and genotype as fixed effects and litter and sire as random effects. The total number of individuals used for SNPs *G31E*, *G26D*, and *K377Q* were 296, 229, and 238, respectively. Pairwise comparisons between genotype least squares means (LSM) were based on the Tukey test.

Single-marker association analyses were also utilized to validate diet-dependent SNP effects. The population was separated into subsets based on diet (standard or energy restricted) to test SNP effects on AP (n = 8 SNPs) and LTNP (n = 4 SNPs) with additive and dominance general linear mixed models including batch, diet, and SNP as fixed effects and litter as a random effect. The entire data set was also analyzed using the same models as well as additive and dominance models that included SNP by diet as a fixed effect in addition to other effects.

Gene Ontology

Regions identified by GWAS that appeared to interact with diet to influence AP or LTNP were extended by 0.5 Mb in both directions for functional characterization of positional candidate genes using the *Sus scrofa* build 10.2. The BIOMART tool in the Ensembl database (http://uswest.ensembl.org/biomart/martview/8aae4d40f586eab6d5336a79044e6bf1) was used to identify positional candidate genes and their gene ontology terms. Functional annotation, gene ontology term enrichment, and pathway analyses were performed on the human orthologs using DAVID (http://david.abcc.ncifcrf.gov).

RESULTS AND DISCUSSION

AVPR1A sequencing and SNP discovery

Sequencing of the *AVPR1A* coding region uncovered three non-synonymous SNPs (Figure 6). The first SNP, *G31E*, is included on the Porcine SNP60 BeadArray, and has been previously shown to be associated with AP and LTNP (Tart *et al.*, 2013). This SNP causes the substitution of a large, negatively charged amino acid (glutamate) for a small, neutral amino acid (glycine) at position 31 in the protein. This SNP is located in the N-terminus of the AVPR1A protein; this region, particularly residues 37-47, was found to be essential for arginine vasopressin and other agonist binding, receptor activation, and second messenger generation (Hawtin *et al.*, 2000). The second SNP, *G256D*, was located in the third intracellular loop, and also results in the substitution of a large, negatively charged amino acid (aspartate) for a small, neutral amino acid (glycine). Finally, the third SNP, *K377Q*, results in an amino acid change from a positive amino acid (lysine) to a neutral amino acid (glutamine) in the C-terminus of the AVPR1A protein. Thibonnier *et al*. (2001) demonstrated that without the proximal end of the C-terminus, internalization and recycling of the receptor was reduced, coupling to phospholipase C was altered, and arginine vasopressin stimulation of DNA synthesis and progression through the cell cycle was prevented.

There was a difference in allelic frequencies of 0.19 between early and late AP individuals for SNPs *G31E* and *G256D*, which appear to be in complete linkage disequilibrium (Table 11). The allelic frequency difference was only 0.06 between early and late AP individuals for SNP *K377Q*. The Meishan individuals sequenced had a higher frequency of the unfavorable allele than the favorable allele for *G31E* and *G256D*. While this is not what we expected based on their superior reproductive performance, reproductive traits are highly polygenic traits,

influenced by many genes with small effects. While these particular alleles may not be favorable, Meishans probably have high frequency of many other favorable alleles for reproductive traits across the genome.

Single-marker association analyses and haplotype effects

Single-marker association analyses revealed a suggestive effect of SNP *G31E* on AP (P = 0.09) with the *11* (*GG*) and *22* (*AA*) genotype LSM being suggestively different (P < 0.10; Table 12). The *G31E* SNP was significantly associated with LTNP (P < 0.05), with genotype *11* (*GG*) and *22* (*AA*) LSM being significantly different (P < 0.05; Table 13). Genotyping confirmed SNPs *G31E* and *G256D* were in complete linkage disequilibrium. As such, it would be expected that P-values would be the same between the two SNPs. However, different numbers of high-quality genotypes used in the analysis for each SNP resulted in differing statistical power between the two SNPs, accounting for differences in significance level. SNP *K377Q* did not have an appreciable effect on either trait (P > 0.1). Genotyping revealed three haplotypes, *AGA*, *AGC*, and *GAA* (Figure 7). Haplotype had a suggestive effect on AP and LTNP (P = 0.14). The *GAA* haplotype is favorable for both traits and is suggestively different from the average of both traits (P < 0.07). The *GAA* haplotype is suggestively different from the *AGC* haplotype for AP (P = 0.08) and LTNP (P = 0.06).

Gene expression analysis

Significant differences in *AVPR1A* expression were detected between the anterior and posterior pituitary, with the posterior pituitary showing a higher level of gene expression (Figure 8). No significant differences in expression were detected between early and late AP individuals (Figures 9 and 10), pre- and post-pubertal individuals (Figures 11 and 12), or *G31E* genotype *AA* and *AG* individuals (Figures 13 and 14). Vasopressin is released from the posterior pituitary, whereas the anterior pituitary does not have functions related to the vasopressin pathway

(Hadley and Levine, 2007). It is likely that expression is higher in the posterior pituitary due to negative feedback mechanisms that presumably influence vasopressin release. On the other hand, *AVPR1A* expression is known to be highest in the hypothalamus, the main location where AVPR1A mediates vasopressin's effects (Hadley and Levine, 2007). While significant differences in expression were not detected in the pituitary, it is likely that significant differences in expression exist in the hypothalamic nuclei that control various aspects of sexual and social behavior. Low numbers of individuals were utilized in these analyses, resulting in low statistical power, particularly when analyzing expression differences between *AVPR1A G31E* genotypes. With a higher number of individuals, it may be possible to detect expression differences that we were unable to find here.

Influence of energy intake during development on age at puberty and reproductive longevity

Energy restriction delayed AP by seven days (Figure 15). Despite this increase in AP, energy restricted gilts were significantly more likely to produce parities two (P = 0.04) and three (P = 0.03). Klindt *et al.* (2001a) reported breeding period compensatory gains in gilts fed 26% energy restricted diets during development that allowed them to achieve adequate body weight to exhibit first estrus at the same time as their contemporaries fed ad libitum throughout the study. As such, conception and farrowing rates were not different between groups. In the study by Klindt *et al.* (2001a), energy restriction was enacted between 90 and 175 days of age, whereas in this study, gilts were fed energy restricted diets between 120 and 240 days of age. We do see increased AP as energy restriction occurs until the start of the breeding period. However, it is likely that compensatory gains occur in energy restricted gilts during gestation so that body weight is not different between energy restricted and non-energy restricted gilts by parity two. Merry (2002) reviews the mechanism believed to explain increased lifespan in rodents fed energy restricted diets; calorie-restricted feeding appears to lower the inner

mitochondrial membrane potential, reducing the rate of free radical generation and, therefore, tissue oxidative damage. This same mechanism may potentially influence reproductive success rates in our study.

Interaction between *AVPR1A G31E* genotype and energy intake

AVPR1A SNP *G31E* interacted with energy intake to suggestively influence the probability to generate parity two (P = 0.07) and significantly affect probability to generate parity three (P = 0.03). The *GG* genotype is favorable under both dietary conditions. However, if fed an energy restricted diet during development, individuals with the *AG* genotype are just as likely to produce parity two as individuals with the *GG* genotype fed a standard, ad libitum diet (Figure 16). Significant differences exist between the *GG* and *AA* genotype (P < 0.01) and *GG* and *AG* genotype (P < 0.05) LSM for probability to generate parity two.

Genome-wide association analysis

The proportion of phenotypic variance of AP explained by markers was 28.26% when only SNP effects were considered. SNP by diet interaction effects did not account for an appreciable amount of phenotypic variance as the proportion of phenotypic variance explained by markers only increased by 0.25% when interaction effects were considered (Table 14). Interaction QTL regions for AP were located on SSC6 (8.5 Mb, two SNPs), SSC7 (4.5 Mb), SSC8 (3.9 Mb), SSC11 (56.5 Mb), SSC14 (51.7 Mb), and SSC16 (14.4 Mb; Figure 17). Candidate genes in these regions include *ACOX3* (SSC8, 4.4 Mb), which plays a role in fatty acid and lipid metabolism, *SPRY2* (SSC11, 56.9 Mb), which has effects on regulation of MAPK and embryonic development, and *SLC5A* (SSC14, 52.0 Mb), whose functions include carbohydrate and sugar uptake and transport as well as intestinal absorption. A QTL SNP with an unmapped location was also identified.

Four genotype by diet interaction QTL regions were identified for LTNP (Figure 18). These QTL were located on SSC4 (31.9 Mb), SSC6 (76.9 Mb), SSC9 (11.6 Mb), and SSC16 (2.4 Mb). The regions nearby these QTL harbor several candidate genes involved in reproductive and/or metabolic processes. SSC4 contains *EIF3E* (SSC4, 31.7 Mb), known to play a role in protein metabolism, and *RSPO2* (SSC4, 31.9 Mb), whose functions include carbohydrate and polysaccharide binding, embryonic limb morphogenesis, and bone mineralization (gene ontology.org). *PAQR7* (SSC6, 76.9 Mb), which is a progesterone receptor and has effects on oogenesis and development, and *PAFAH2* (SSC6, 77.1 Mb), which is involved in lipid metabolism (geneontology.org), are located on SSC6. Finally, SSC9 is home to *MOGAT2* (SSC9, 11.1 Mb), known to play a role in lipid metabolism, synthesis, and absorption and may also be involved in diet-induced obesity (geneontology.org). No common diet-dependent QTL regions existed between AP and LTNP.

Single-marker association analyses

When the entire data set was analyzed, none of the genotype by diet interaction QTL SNPs had significant additive effects on AP ($P > 0.05$). However, all eight SNPs had a significant genotype by diet interaction effect on AP under an additive model ($P < 0.05$). When the dataset was separated by diet, all SNPs had a significant effect in one diet or both (Table 15). These SNPs only exhibited significant effects in one diet or had opposite effects in each dietary treatment. For example, the *CC* genotype of *MARC0053591* was favorable in gilts fed a standard diet, but the *TT* genotype was favorable in gilts fed an energy-restricted diet (Figure 19). This SNP had significant additive ($P < 0.0001$) and dominance ($P < 0.05$) effects on AP when only gilts fed the standard diet were considered and significant additive effects ($P < 0.01$) when only gilts fed the energy restricted diet were considered. The *TT* genotype LSM was significantly different from both the *CT* and *CC* LSM ($P < 0.05$) in gilts fed the standard diet, and the *CT* and *CC* genotype

LSM differ ($P < 0.05$) in gilts fed the energy restricted diet. When all gilts were included in the analyses without accounting for interaction effects, no significant effects on AP were observed.

The genotype by diet interaction QTL SNPs for LTNP behaved similarly. All four SNPs had a significant additive genotype by diet interaction effect ($P < 0.05$), but none had an effect on LTNP alone when all gilts were considered in the analysis ($P > 0.05$). All SNPs had a significant additive influence ($P < 0.05$) on LTNP when the SNPs were analyzed separately in the standard and energy restricted diets (Table 16).

CONCLUSION

Improving sow reproductive longevity is very important to the swine industry as reproductive failure is the most common reason for early culling and leads to reduced sow productivity and higher replacement costs. Marker-assisted selection may be necessary to overcome obstacles faced when using traditional selection to improve reproductive longevity. Age at puberty is a correlated trait that can be utilized to identify markers which also influence reproductive longevity. Genome-wide association analysis uncovered a region on SSC5 which influences both AP and LTNP. One candidate gene located in this region, *AVPR1A*, appears to have significant effects on both traits. Two novel non-synonymous SNPs were identified in addition to a previously known SNP. SNP *G31E* causes a non-conservative amino acid substitution near a region essential for agonist binding, receptor activation, and second messenger generation. SNP *K377Q* results in another non-conservative amino acid substitution near a region that influences downstream signaling. SNPs *G31E* and *G256D* were in complete linkage disequilibrium and were significantly associated with LTNP and suggestively associated with AP. Three haplotypes were identified, and the *GAA* haplotype was suggestively associated with younger AP and increased LTNP compared to the *AGC* haplotype. No major differences in *AVPR1A* expression were observed between individuals expressing first estrus early and individuals expressing first estrus late nor between individuals with the *G31E AG* genotype and individuals with the *G31E AA* genotype. Selection based on SNPs such as *G31E* and *G256D* have the potential to reduce AP and improve reproductive longevity. This will lead to an increase in sow net values in commercial herds.

Previous studies have shown that caloric restriction increases lifespan in vertebrate species, and no adverse effects on reproduction were observed in gilts fed reduced energy diets

during growth and development. However, no studies had analyzed interactions between developmental energy intake and genetic markers on a genome-wide level on traits such as age at puberty and reproductive longevity. Reduced energy intake results in later expression of first estrus and increased probability to generate parities two and three. *AVPR1A G31E* genotype interacted with diet to significantly influence probability to generate parity three. Interaction effects exist between genotypes and energy intake which influence AP and LTNP, though these SNPs explain a very small proportion of the phenotypic variance for these traits. Candidate genes are located in these regions which perform various functions including energy metabolism, nutrient absorption, gamete generation, and embryonic development. SNPs identified as interacting with energy intake during development to influence AP and LTNP only had an appreciable effect in individuals fed one dietary treatment or the direction of the effect was opposite between the two dietary treatments. While these markers may not be very useful to producers when making selection decisions due to their small effects that are dependent on developmental energy intake, utilizing lower energy ingredients in feed or restricting intake during development may provide a cost-saving benefit while maintaining or even improving herd reproductive performance.

LITERATURE CITED

Bidanel, J.P. 2011. Biology and genetics of reproduction. In *The Genetics of the Pig.* 2nd Edition. Edited by Rothschild, M.F. and Ruvinsky, A. Wallingford, UK: CAB International; 218-241.

DAVID Functional Annotation Bioinformatics Microarray Analysis http://david.abcc.ncifcrf.gov.

Ensembl Genome Browser BioMart http://uswest.ensembl.org/biomart/martview/8aae4d40f586eab6d5336a79044e6bf1.

Fernando, R.L. and Garrick, D.J. 2008. Gensel User Manual for Portfolio of Genomic Selection Related Analyses. Animal Breeding and Genetics. Iowa State University, Ames.

Gene Ontology Consortium http://geneontology.org.

Hadley, M.E. and Levine, J.E. 2007. Endocrinology. 6th Edition. Pearson.

Hammock, E.A.D. and Young, L.J. 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. Science. 308:1630-1634.

Hawtin, S.R., Wesley, V.J., Parslow, R.A., Patel, S., and Wheatley, M. 2000. Critical role of a subdomain of the N-terminus of the V1a vasopressin receptor for binding agonists but not antagonists; functional rescue by the oxytocin receptor N-terminus. Biochemistry. 39:13524-13533.

Hsu, W.L. 2011. Analysis of long-term selection (28 generations) for reproduction, growth, and carcass traits in swine. PhD dissertation. University of Nebraska – Lincoln, Animal Science Department.

Klindt, J., Yen, J.T., and Christenson, R.K. 1999. Effect of prepubertal feeding regimen on reproductive development of gilts. Journal of Animal Science. 77:1968-1976.

Klindt, J., Yen, J.T., and Christenson, R.K. 2001a. Effect of pre-pubertal feeding regimen on reproductive development and performance of gilts through the first parity. Journal of Animal Science. 79:787-795.

Knauer, M.T., Cassady, J.P., Newcom, D.W., and See, M.T. 2011. Phenotypic and genetic correlations between gilt estrus, puberty, growth, composition, and structural conformation traits with first-litter reproductive measures. Journal of Animal Science. 89:935-942.

Merry, B.J. 2002. Molecular mechanisms linking calorie restriction and longevity. International Journal of Biochemistry and Cell Biology. 34:1340-1354.

Miller, P.S., Moreno, R., and Johnson, R.K. 2011. Effects of restricting energy during the gilt developmental period on growth and reproduction of lines differing in lean growth rate: Responses in feed intake, growth, and age at puberty. Journal of Animal Science. 89:342-354.

Mote, B.E., Koehler, K.J., Mabry, J.W., Stalder, K.J., and Rothschild, M.F. 2009. Identification of genetic markers for productive life in commercial sows. Journal of Animal Science. 87:2187-2195.

NRC. 1998. Nutrient requirements of swine. 10th ed. Natl. Acad. Press, Washington, DC.

Patterson, J.L., Beltranema, E., and Foxcroft, G.R. 2010. The effect of gilt age at first estrus and breeding on third estrus on sow body weight changes and long-term reproductive performance. Journal of Animal Science. 88:2500-2513.

Serenius, T. and Stalder, K.J. 2006. Selection for sow longevity. Journal of Animal Science. 84(Suppl):E166-171.

Simon, P. 2003. Q-Gene: processing quantitative real-time RT-PCR data. Bioinformatics. 19(11):1439-1440.

Stalder, K.J., Lacy, R.C., Cross, T.L., and Conaster, G.E. 2003. Financial impact of average parity of culled females in a breed-to-wean swine operation using replacement gilt net present value analysis. Journal of Swine Health in Production. 11:69-74.

Stalder, K.J. 2004. Sow longevity scrutinized. National Hog Farmer. Primedia Magazines and Media, Inc., Overland Park, KS. 49:26-30.

Tart, J.K., Johnson, R.K., Bundy, J.W., Ferdinand, N.N., McKnite, A.M., Wood, J.R., Miller, P.S., Rothschild, M.F., Spangler, M.L., Garrick, D.J., Kachman, S.D., and Ciobanu, D.C. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Animal Genetics. 44:387-397.

Thibonnier, M., Plesnicher, C.L., Berrada, K., and Berti-Mattera, L. 2001. Role of the human V_1 vasopressin receptor COOH terminus in internalization and mitogenic signal transduction. Am. J. Physiol. Endocrinol. Metab. 281:E81-E92.

Walum, H., Westberg, L., Henningsson, S., Neiderhiser, J.M., Reiss, D., Igl, W., Ganiban, J.M., Spotts, E.L., Pedersen, N.L., Eriksson, E., and Lichtenstein, P. 2008. Genetic variation in the vasopressin receptor 1a gene (AVPR1A) associates with pair-bonding behavior in humans. PNAS. 105(37):14153-14156.

TABLES

Table 1 - Proportion of polymorphic SNPs, inbreeding coefficient (F_{IS}), and average proportion of observed and expected heterozygosity per population

Population	Polymorphic SNPs, %	F_{IS}	Observed Heterozygosity, %	Expected Heterozygosity, %
NIL	73.4	-0.0467	26.4	25.1
CTRL	76.2	-0.0396	27.5	26.3
DxH	83.3	-0.0642	30.0	28.2

Table 2 - Posterior means of variance components of litter size traits based on 56,424 SNP effects estimated by Bayes B

Trait*	n	Genetic variance	Residual variance	Total variance	Phenotypic variance explained by SNPs, %
NBA-P1	903	1.04	12.63	13.67	7.6
NBA-P2	903	0.30	11.60	11.90	2.5
TNB-P1	903	0.36	9.29	9.65	3.7
TNB-P2	903	0.38	12.04	12.42	3.1

*NBA-P1, number born alive at parity 1; NBA-P2, number born alive at parity 2; TNB-P1, total number born at parity 1; TNB-P2, total number born at parity 2.

Table 3 - Major 1-Mb regions explaining phenotypic variation of litter size traits

	NBA-P1*	NBA-P2*	TNB-P1*	TNB-P2*
Rank§	Position Proportion£ P > 0¶	Position Proportion£ P > 0¶	Position Proportion£ P > 0¶	Position Proportion£ P >0¶
1	SSC9, 4-5 Mb 0.95 0.32	SSC6, 70-71 Mb 0.18 0.30	SSC14, 11-12 Mb 0.14 0.35	SSC2, 12-13 Mb 0.13 0.32
2	SSC7, 15-16 Mb 0.30 0.24	SSC14, 19-20 Mb 0.14 0.26	SSC2, 89-90 Mb 0.13 0.26	SSC10, 64-65 Mb 0.12 0.25
3	SSC13, 36-37 Mb 0.26 0.27	SSC1, 53-54 Mb 0.12 0.21	SSC19, 22-23 Mb 0.12 0.18	SSC14, 50-51 Mb 0.12 0.24

*NBA-P1, number born alive at parity 1; NBA-P2, number born alive at parity 2; TNB-P1, total number born at parity 1; TNB-P2, total number born at parity 2.
§Window rank by proportion of genetic variance explained.
£Proportion of genetic variance explained by the window.
¶Probability that window effects are greater than zero.

Table 4 – Multiple trait QTL regions for litter size traits

Region	Traits
SSC1 34-35 Mb	NBA-P1 NBA-P2 TNB-P2
SSC2 12-13 Mb	NBA-P2 TNB-P1 TNB-P2
SSC6 93-94 Mb	NBA-P2 TNB-P2
SSC9 147-148 Mb	NBA-P2 TNB-P1 TNB-P2
SSC13 36-37 Mb	NBA-P1 TNB-P1
SSC14 19-20 Mb	NBA-P2 TNB-P2

Table 5 – Significant core haplotypes, candidate genes, and their functional role

Core Haplotype Region	Candidate Gene	Function	Associated Disorders
SSC2 6.9-7.7 Mb	*MEN1* (6.5 Mb) *SF1* (6.5 Mb) *ESRRA* 6.9 Mb)	Pregnancy, embryonic development, embryogenesis Sex differentiation Estrogen receptor	Premature Ovarian Failure
SSC3 133.2-133.7 Mb	N/A		
SSC3 133.9-134.1 Mb	*ROCK2* (134.2 Mb)	Embryogenesis, neural tube closure	
SSC4 11.2-12.6 Mb	*MYC* (12.8 Mb)	Embryonic organ development	
SSC5 10.7-11.0 Mb	N/A		
SSC6 8.7-8.9 Mb	N/A		
SSC6 9.0-9.3 Mb	N/A		
SSC6 34.4-35.4 Mb	*CCNE1* (35.2 Mb)	Ovarian follicle development & growth, sex differentiation, response to estrogen & progesterone, gonad development	
SSC6 36.2-36.7 Mb	*ZNF536* (36.0 Mb)	Embryonic development	
SSC7 13.2-13.8 Mb	*ATXN1* (12.9 Mb)	Learning, memory	Neurodegeneration
SSC8 5.2-5.5 Mb	*MSX1* (5.2 Mb) *DRD5* (5.7 Mb) *OTOP1* (5.7 Mb)	Embryonic development, activation of meiosis Mating behavior, regulation of female receptivity, learning Embryonic development	
SSC9 15.4-15.8 Mb	N/A		
SSC9 15.9-17.5 Mb	N/A		
SSC9 47.6-48.4 Mb	*CADM1* (47.7 Mb)	Immune response	
SSC9 53.7-55.0 Mb	*TECTA* (53.5 Mb) *CRTAM* (55.1 Mb)	Neurological system process, cognition Immune response	

Table 5 – Significant core haplotypes, candidate genes, and their functional role, continued

Core Haplotype Region	Candidate Gene	Function	Associated Disorders
SSC9 138.2-139.0 Mb	*RNF2* (138.7 Mb)	Embryonic morphogenesis, gastrulation, germ cell development	
SSC10 35.5-36.1 Mb	*IL11RA* (36.3 Mb) *ARID3C* (36.3 Mb) *CNTFR* (36.4 Mb)	Embryo implantation, pregnancy Embryonic patterning Sex differentiation, brainstem development	
SSC11 25.0-25.3 Mb	*TNSF11* (25.1 Mb)	Immune system development, immune response	
SSC12 10.7-11.1 Mb	N/A		
SSC12 42.5-42.9 Mb	*CCL1* (42.5 Mb) *CCL2* (42.5 Mb)	Immune response Response to progesterone, reproduction, immune response	
SSC12 58.4-59.4 Mb	*MYH13* (57.9 Mb) *MYH1* (58.0 Mb)	Immune response Immune response	
SSC15 152.8-155.8 Mb	*TRAF3IP1* (152.4 Mb) *TWIST2* (152.6 Mb) *HDAC4* (152.9 Mb)	Embryonic development, neural tube patterning Embryonic morphogenesis Immune system development, immune response	
SSC16 79.5-79.7 Mb	*MTRR* (80.3 Mb)		Neural tube defects
SSC16 80.5-81.3 Mb	*ADCY2* (80.4-80.6 Mb) *SRD5A1* (81.7 Mb)	Oocyte maturation & meiosis, GnRH signaling pathway Sex differentiation, androgen & estrogen metabolism	PCOS (Tsilchorozidou *et al*., 2003)
SSCX 2.1-3.2 Mb	*NLGN4X* (2.8-3.2 Mb)	Learning, brainstem & cerebellum development	Autism and Asperger Syndrome
SSCX 6.8-7.6 Mb	*TBL1Y* (6.5 Mb) *GPR143* (6.6 Mb) *SHROOM2* (6.6-6.8 Mb)	Response to estrogen Neurological system process, cognition Brain development	

Table 6 – Major regions with significantly different allelic frequencies between populations identified via contingency tests

Region	Allelic Frequency Differences	QTL	Candidate Gene	Function
SSC3 114-115 Mb	NIL vs. CTRL	NBA-P1	*SRD5A2* (114.6 Mb)	sex differentiation
			XDH (114.8 Mb)	lactation, mammary gland development, and reproduction
SSC12 21-22 Mb	NIL vs. DxH	NBA-P1	*STAT5A* (20.7 Mb)	sex differentiation, gonad development, ovulation cycle, response to estradiol, progesterone metabolism, pregnancy
			STAT3 (20.7 Mb)	sexual reproduction and response to estrogen stimulus
			STAT5B (20.8 Mb)	sex differentiation, gonad development, ovulation cycle, response to estradiol, progesterone metabolism, pregnancy
			KAT2A (20.9 Mb)	embryonic development, including neural tube closure
			KRT19 (21.5 Mb)	estrogen stimulus, cell differentiation during embryonic placenta development
			TOP2A (22.3 Mb)	embryonic cleavage
			RARA (22.4 Mb)	germ cell development, estrogen receptor signaling pathway
			THRA (22.5 Mb)	female courtship behavior and receptivity
SSC13 32-33 Mb	NIL vs. CTRL NIL vs. DxH	TNB-P1	*TGDF1* (32.8 Mb)	in utero embryonic development, gastrulation, embryo anterior/posterior axis specification
			SETD2 (33.1 Mb)	embryonic morphogenesis and development, neural tube closure
SSCX 22-23 Mb	NIL vs. DxH	TNB-P1	*ZFX* (21.6 Mb)	fertilization, oocyte, ovarian follicle, embryo, and germ cell development, parental behavior
			PCYT1B (22.1 Mb)	sex differentiation, development of gonads and ovarian follicles, gamete generation, the ovulation cycle

Table 7 – Major regions with significantly different allelic frequencies between NIL and CTRL identified via Wright's fixation index (F_{ST})

Region	Number 99.9% SNPs	Other Methods*	Candidate Gene	Function
SSC1 129-131 Mb	1	Heterozygosity Reduction	*NEDD4* (128.5 Mb)	Progesterone receptor signaling
			DYX1C1 (128.9 Mb)	Estrogen recptor binding and signaling
SSC2 13-14 Mb	2	QTL (SSC2, 12-13 Mb)	*GLRX* paralog (12.1 Mb)	Implantation (Stavréus-Evers *et al.*, 2002)
			P2X3R (13.4 Mb)	Behavior/neurological, immune system
			SSRP1 (13.4 Mb)	Embryogenesis
SSC2 43-45 Mb	4		*E2F8* (43.2 Mb)	Placenta development, trophoblast giant cell differentiation
SSC2 124-129 Mb	10	NIL vs. CTRL contingency	*TRIM36* (124.1 Mb)	Sexual reproduction
SSC4 9-10 Mb	1	QTL	*LRRC6* (8.6 Mb)	Reproductive system development
SSC4 109-111 Mb	11	NIL vs. CTRL contingency	*TXNIP* (109.0 Mb)	Response to estrogen and progesterone stimulus
			NOTCH2 (110.9 Mb)	Placenta and in utero embryonic development
			ADAM30 (111.2 Mb)	Binding of sperm to the zona pellucida, single fertilization
SSC8 98-99 Mb	1	NIL vs. CTRL contingency		

* Other methods in which this region was identified as important. QTL – explained phenotypic variation for multiple litter size traits or a large proportion of the variation for one trait. Heterozygosity reduction – exhibited ≥ four-fold reduction in heterozygosity in NIL compared to CTRL. NIL vs. CTRL contingency – contained multiple SNPs with significantly different allelic frequencies between NIL and CTRL. REHH (Relative Extended Haplotype Homozygosity) – identified as a significant core haplotype likely under selection in NIL

Table 7 – Major regions with significantly different allelic frequencies between NIL and CTRL identified via Wright's fixation index (F_{ST}), continued

Region	Number 99.9% SNPs	Other Methods*	Candidate Gene	Function
SSC10 56-57 Mb	**6**		*KIAA1217* (56.2 Mb)	Embryonic development
			PTF1A (57.5 Mb)	Embryonic development
SSC11 78-79 Mb	2		*FGF14* (77.7 Mb)	Embryonic development, defense response to virus
SSC12 59-60 Mb	1	REHH (SSC12 58.4-59.4 Mb), QTL (SSC12, 60-61 Mb)		
SSC15 135-136 Mb	1	QTL	*INHA* (134.5 Mb)	Follicle development, ovulation cycle, GnRH and FSH secretion
SSC17 47-49 Mb	**3**		*BPI* (46.8 Mb)	Immune response
			LBP (46.8 Mb)	Immune response
			MAFB (48.7 Mb)	Embryonic organ development
			TOP1 (49.1 Mb)	Embryonic cleavage
			PLCG1 (49.2 Mb)	Embryonic development
SSCX 106-107 Mb	1	NIL vs. CTRL contingency	*AMOT* (107.3 Mb)	In utero embryonic development

* Other methods in which this region was identified as important. QTL – explained phenotypic variation for multiple litter size traits or a large proportion of the variation for one trait. Heterozygosity reduction – exhibited ≥ four-fold reduction in heterozygosity in NIL compared to CTRL. NIL vs. CTRL contingency – contained multiple SNPs with significantly different allelic frequencies between NIL and CTRL. REHH (Relative Extended Haplotype Homozygosity) – identified as a significant core haplotype likely under selection in NIL

Table 8 – Single Marker Association Analyses for SSC2 QTL and F_{ST} SNPs

SNP	Trait	Additive P	Dominance P	Genotype	LSM (SE)*
ALGA0118548	NBA-P1	0.4203	0.9328	AA	12.65 (0.38)
				AC	12.88 (0.21)
				CC	13.02 (0.33)
	TNB-P1	0.2159	0.8605	AA	13.67 (0.38)
				AC	14.03 (0.22)
				CC	14.23 (0.33)
	NBA-P2	0.2397	0.4486	AA	12.8 (0.51)
				AC	12.03 (0.24)
				CC	11.99 (0.41)
	TNB-P2	0.4786	0.3246	AA	13.77 (0.52)
				AC	13.06 (0.23)
				CC	13.25 (0.41)
ASGA0095946	NBA-P1	0.639	0.6774	GG	12.83 (0.26)
				GT	12.82 (0.21)
				TT	13.04 (0.33)
	TNB-P1	0.496	0.4819	GG	14.04 (0.26)
				GT	14 (0.21)
				TT	14.35 (0.33)
	NBA-P2	0.0798	0.3543	GG	12.82 (0.33)
				GT	12.14 (0.25)
				TT	11.9 (0.44)
	TNB-P2	0.1023	0.6008	GG	13.86 (0.33)
				GT	13.34 (0.26)
				TT	12.98 (0.44)

*LSM – least squares means; SE – standard error

Table 8 – Single Marker Association Analyses for SSC2 QTL and F_{ST} SNPs, continued

SNP	Trait	Additive P	Dominance P	Genotype	LSM (SE)*
ALGA0111847	NBA-P1	0.6882	0.6887	GG	12.93 (0.19)
				GT	12.81 (0.25)
				TT	12.81 (0.78)
	TNB-P1	0.309	0.4166	GG	14.17 (0.19)
				GT	13.91 (0.25)
				TT	13.68 (0.79)
	NBA-P2	0.5484	0.5032	GG	12.3 (0.23)
				GT	12.03 (0.31)
				TT	12.16 (1.02)
	TNB-P2	0.4431	0.6367	GG	13.41 (0.24)
				GT	13.19 (0.32)
				TT	12.75 (1.03)
ASGA0097301	NBA-P1	0.5464	0.5217	CC	12.85 (0.23)
				CT	12.65 (0.26)
				TT	12.69 (0.63)
	TNB-P1	0.0325	0.1279	CC	14.21 (0.22)
				CT	13.66 (0.25)
				TT	13.26 (0.62)
	NBA-P2	0.4294	0.2975	CC	12.48 (0.27)
				CT	12.06 (0.31)
				TT	12.37 (0.85)
	TNB-P2	0.1853	0.1233	CC	13.67 (0.28)
				CT	13.02 (0.32)
				TT	13.31 (0.85)

*LSM – least squares means; SE – standard error

Table 9 – Single-Marker Association Analysis between *P2X3R* SNPs and TNB-P2.

SNP	Model	P	Percent Variance Explained¶	Genotype	N	TNB-P2 LSM (SE)*
P2X3R exon 4	Additive, including AA	0.1695	1.8	GG	60	13.15 (0.99)
				AG	46	11.71 (1.11)
				AA	5	9.72 (3.11)
	Additive, excluding AA	0.3289		GG	60	13.12 (0.87)
				AG	46	11.77 (1.03)
P2X3R exon 7	Additive, including TT	0.0696	4.0	CC	70	13.13 (0.95)
				CT	46	11.68 (1.10)
				TT	4	5.55 (3.47)
	Additive, excluding TT	0.4326		CC	70	13.01 (0.94)
				CT	46	11.86 (1.09)

¶Percent variance explained calculated with additive SNP effects and allelic frequencies (Falconer and Mackay, 1996) expressed as a percent of the total variance.

*LSM – least squares means; SE – standard error

Table 10 – Genotype by diet interactions GWAS SNP effect coding

Diet	SNP Genotype	Main Effects Coded As	Interaction Effects Coded As
Standard	11	-10	-10
	12	0	0
	22	10	10
Energy Restricted	11	-10	10
	12	0	0
	22	10	-10

Table 11 – *AVPR1A* sequencing allelic frequencies

G31E				*G256D*				*K377Q*			
Allele	Early AP Freq	Late AP Freq	Meishan Freq	Allele	Early AP Freq	Late AP Freq	Meishan Freq	Allele	Early AP Freq	Late AP Freq	Meishan Freq
G	0.44	0.25	0.125	A	0.44	0.25	0.125	A	0.69	0.625	1
A	0.56	0.75	0.875	G	0.56	0.75	0.875	C	0.31	0.375	0

Table 12 – *AVPR1A* non-synonymous SNP single marker association analysis for age at puberty

		Genotype Least Squares Means			
	n	11	12	22	Overall P
G31E	296	182.5[a]	176.8	172.0[a]	0.09
G256D	229	170.2	177.8	180.9	0.13
K377Q	238	176.4	177.7	181.5	0.81

a: P < 0.10 Differences between genotype least squares means

Table 13 – *AVPR1A* non-synonymous SNP single marker association analysis for lifetime number of parities

		Genotype Least Squares Means			
	n	11	12	22	Overall P
G31E	296	1.57[a]	1.81	2.30[a]	0.05
G256D	229	2.34	1.74	1.68	0.10
K377Q	238	2.05	1.77	1.05	0.15

a: P < 0.05 Differences between genotype least squares means

Table 14 – Posterior means of variance components of AP based on SNP effects estimated by Bayes B analysis

GWAS	Genetic Variance	Residual Variance	Total Variance	Phenotypic Variance Explained by Markers (%)
SNP	90.3	229.3	319.6	28.26
SNP x Diet	97.3	244.6	342.1	28.51

Table 15 – Number of SNPs with significant effects on AP ($P < 0.05$)

Model	Standard Diet	Energy-Restricted Diet	All Diets Combined	SNP x Diet Interactions
Additive	6	7	0	8
Dominance	1	4	2	5

n = 8

Table 16 – Number of SNPs with significant effects on LTNP ($P < 0.05$)

Model	Standard Diet	Energy-Restricted Diet	All Diets Combined	SNP x Diet Interactions
Additive	3	2	0	4
Dominance	2	1	0	1

n = 4

FIGURES

Figure 1 - Multidimensional scaling of dimensions 1 vs. 2 (top) and dimensions 1 vs. 3 (bottom) for 21,524 SNPs

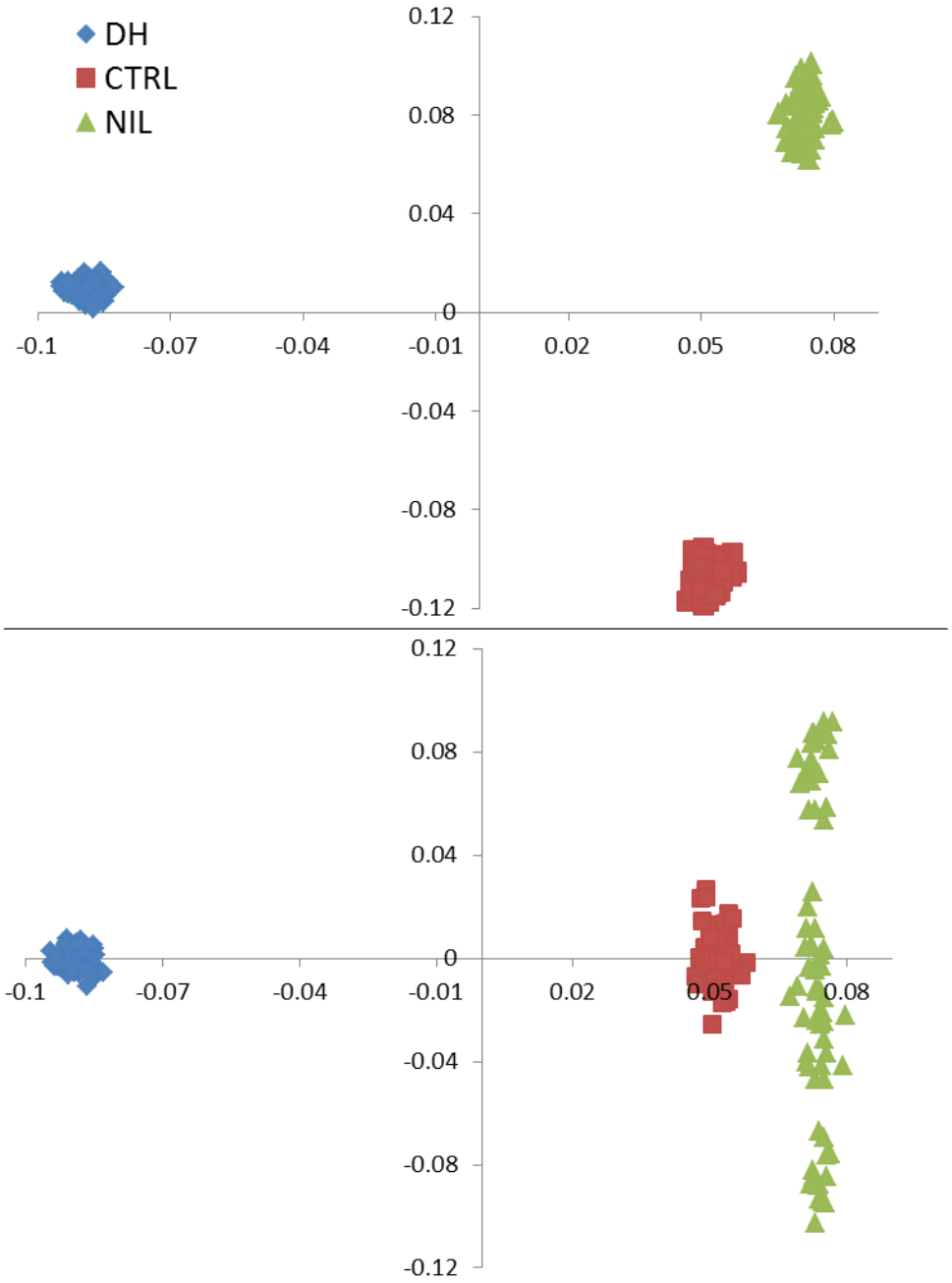
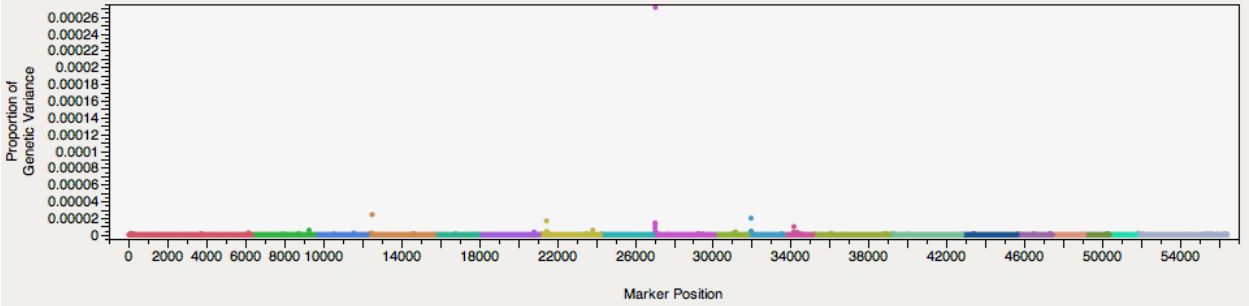


Figure 2 – Proportion of Genetic Variance Explained by Markers, NBA-P1



Each dot represents the genetic variance of NBA-P1 explained by each SNP. The location of each SNP is shown on the x-axis, with each chromosome depicted by a different color. The proportion of genetic variance explained is located on the y-axis.

Figure 3 – Distribution of Observed F_{ST}

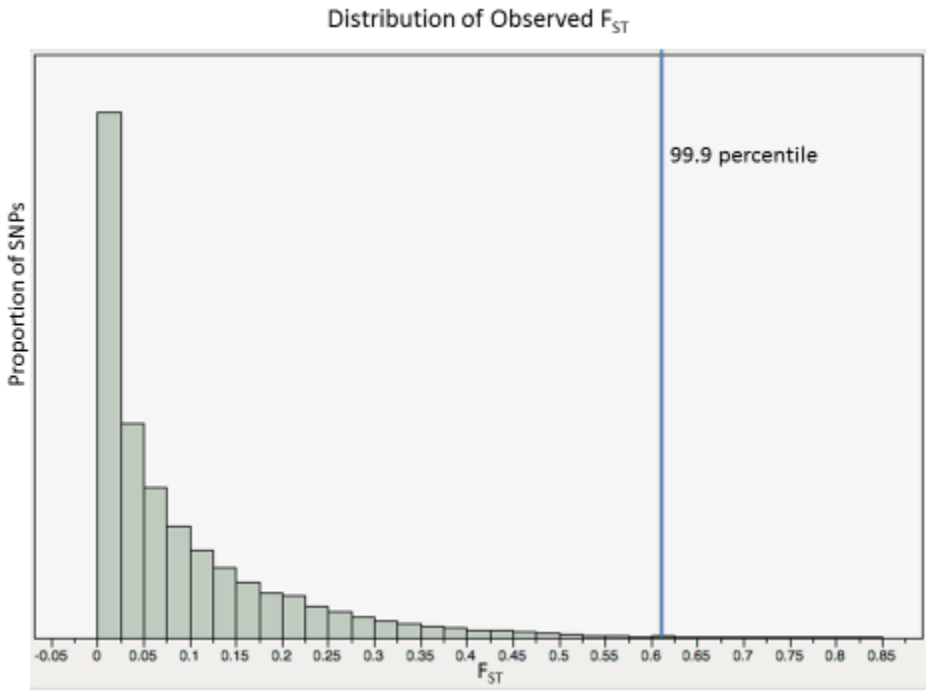


Figure 4 – TNB-P2 least squares means by genotype of *P2X3R* exon 4 SNP

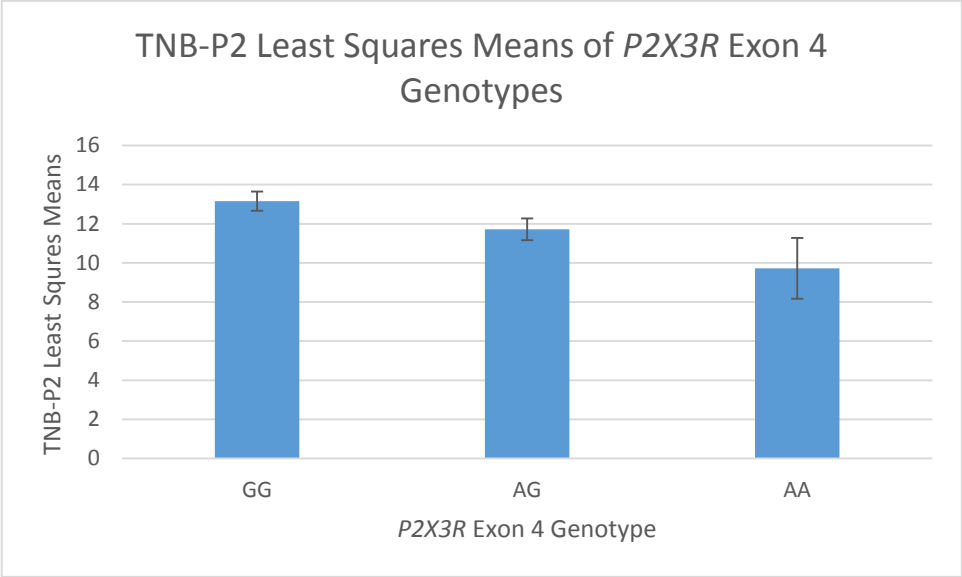


Figure 5 – TNB-P2 least squares means by genotype of *P2X3R* exon 7 SNP

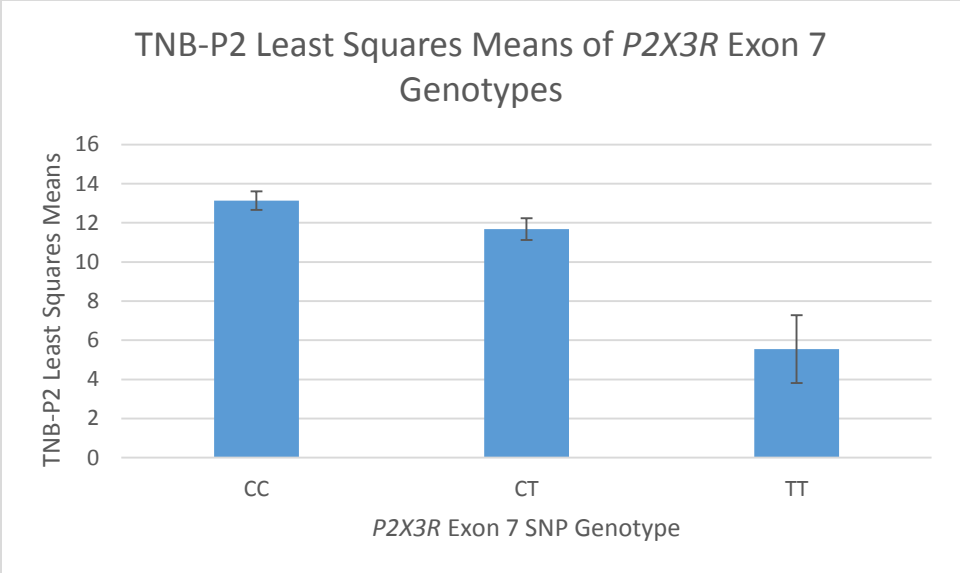


Figure 6 – Location of *AVPR1A* amino acid substitutions

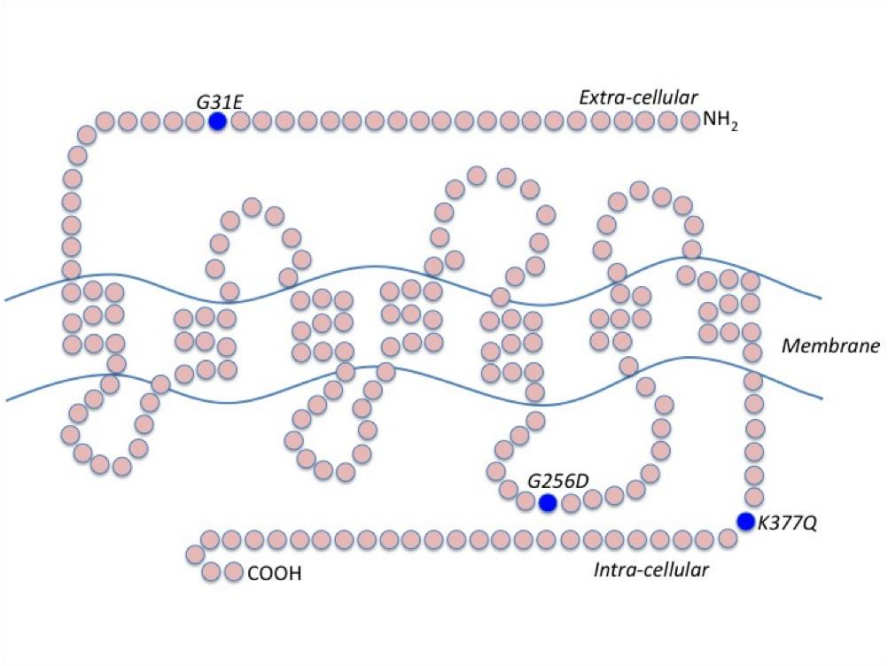
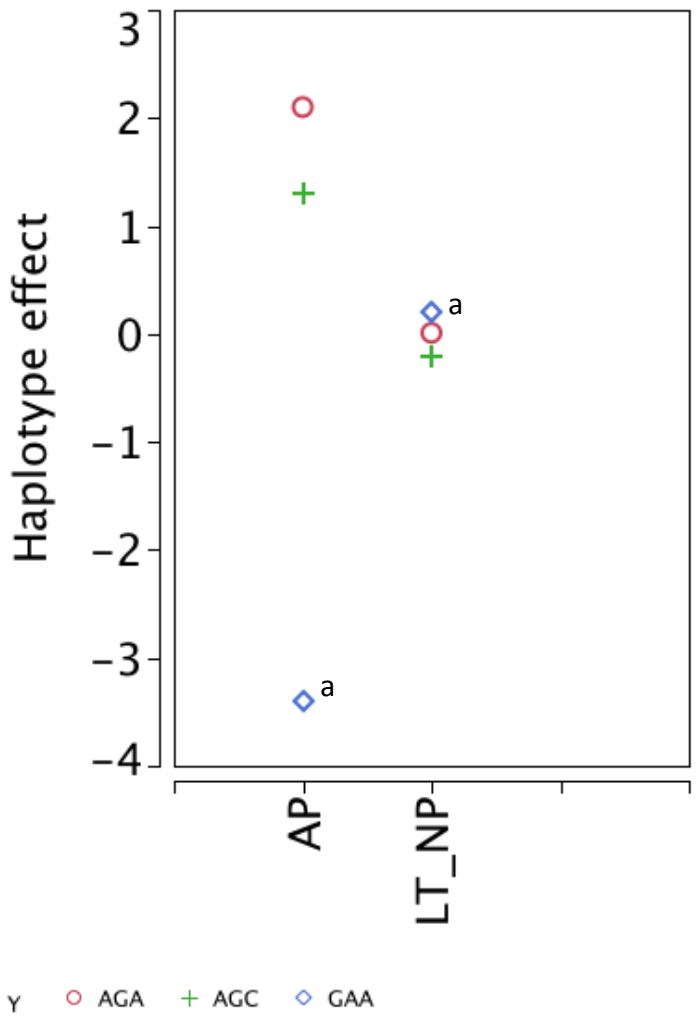


Figure 7 – *AVPR1A* haplotype effects



a: P < 0.07

Figure 8 - *AVPR1A* gene expression: anterior (n = 11) vs. posterior (n = 9) pituitary

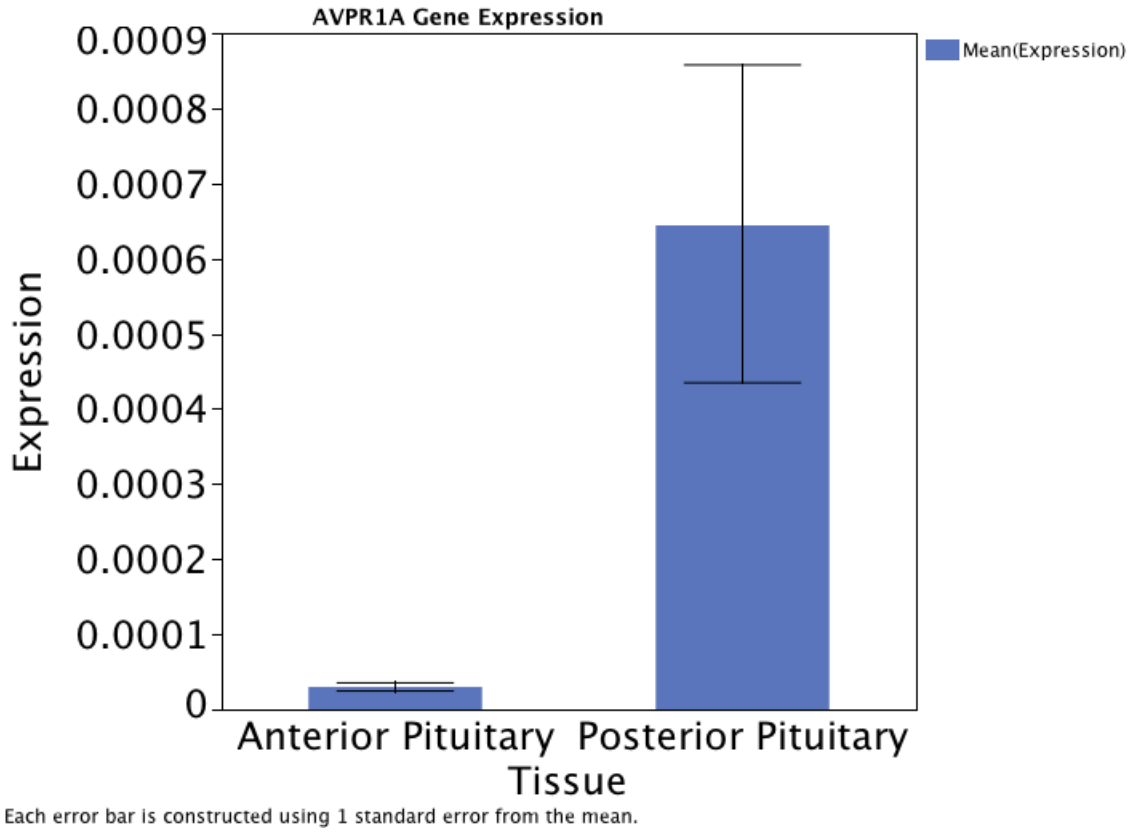


Figure 9 - *AVPR1A* gene expression: early (n = 7) vs. late (n = 3) AP in the anterior pituitary

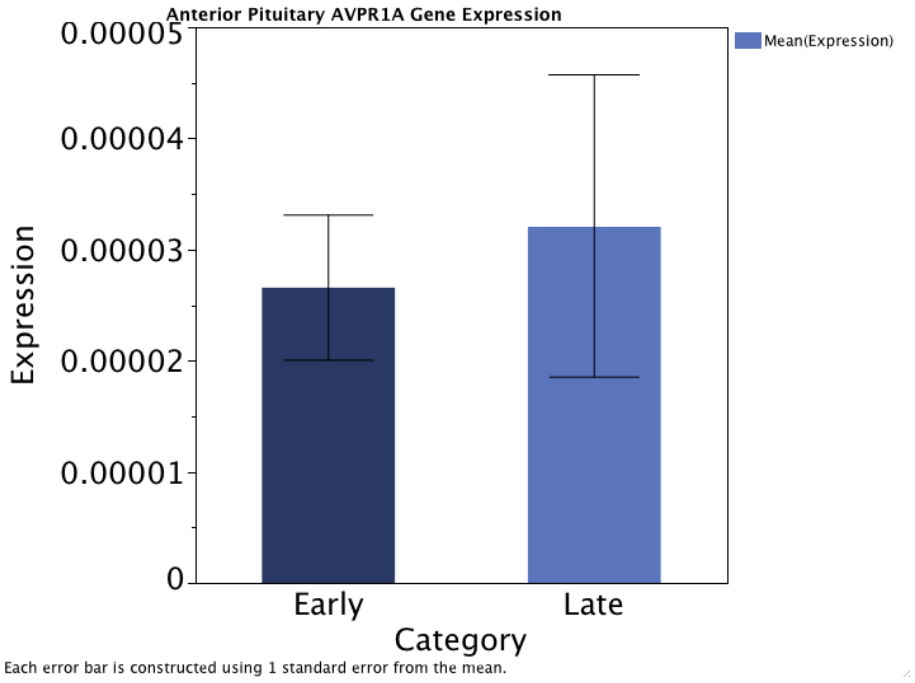


Figure 10 - *AVPR1A* gene expression: early (n = 6) vs. late (n = 2) AP in the posterior pituitary

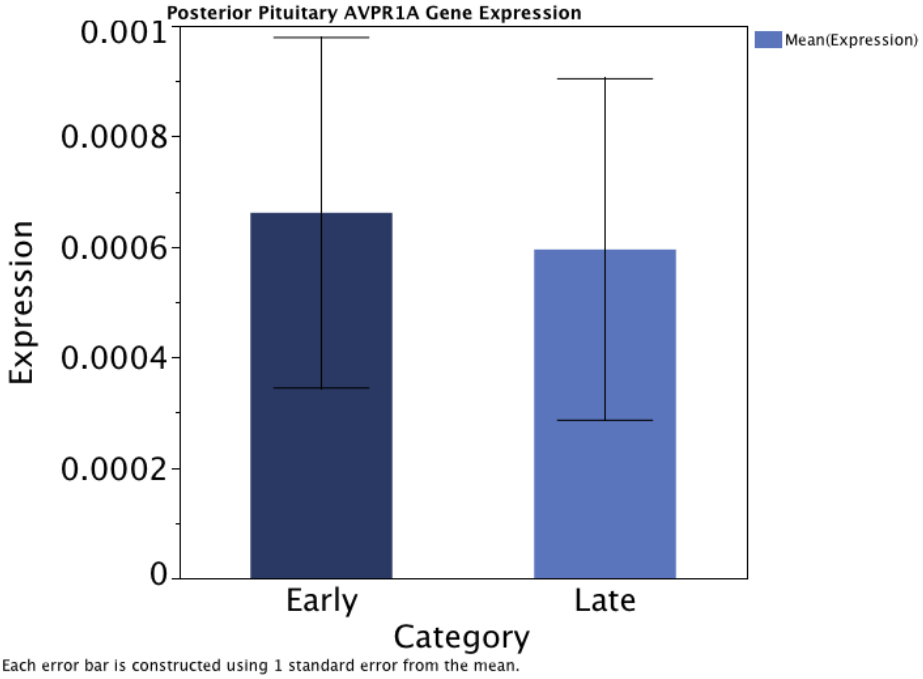


Figure 11 - *AVPR1A* gene expression: pre (n = 9) vs. post (n = 11) pubertal in the anterior pituitary

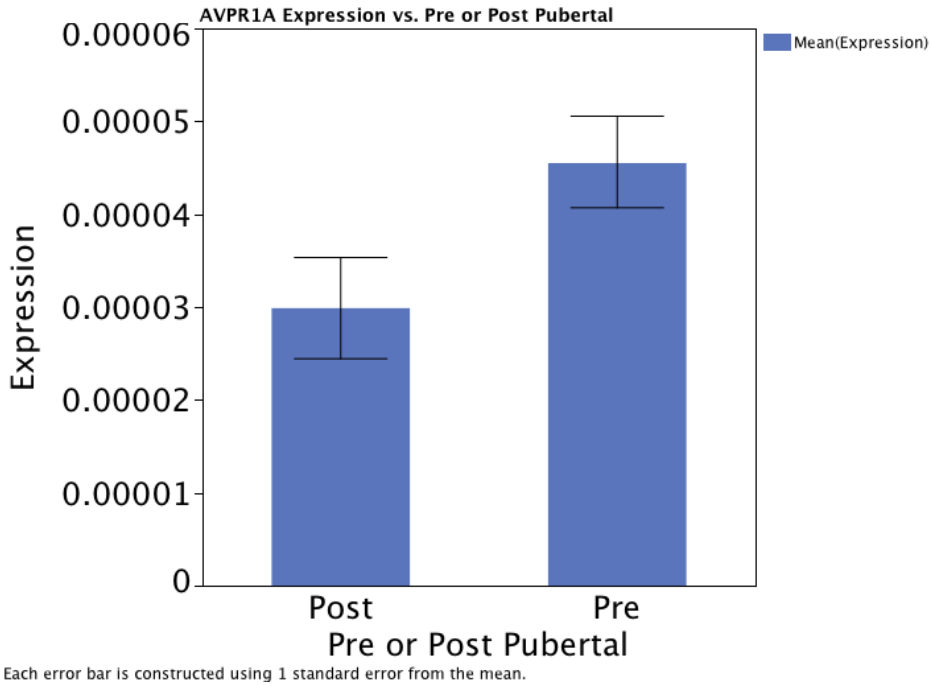


Figure 12 - *AVPR1A* gene expression: pre (n = 8) vs. post (n = 9) pubertal in the posterior pituitary

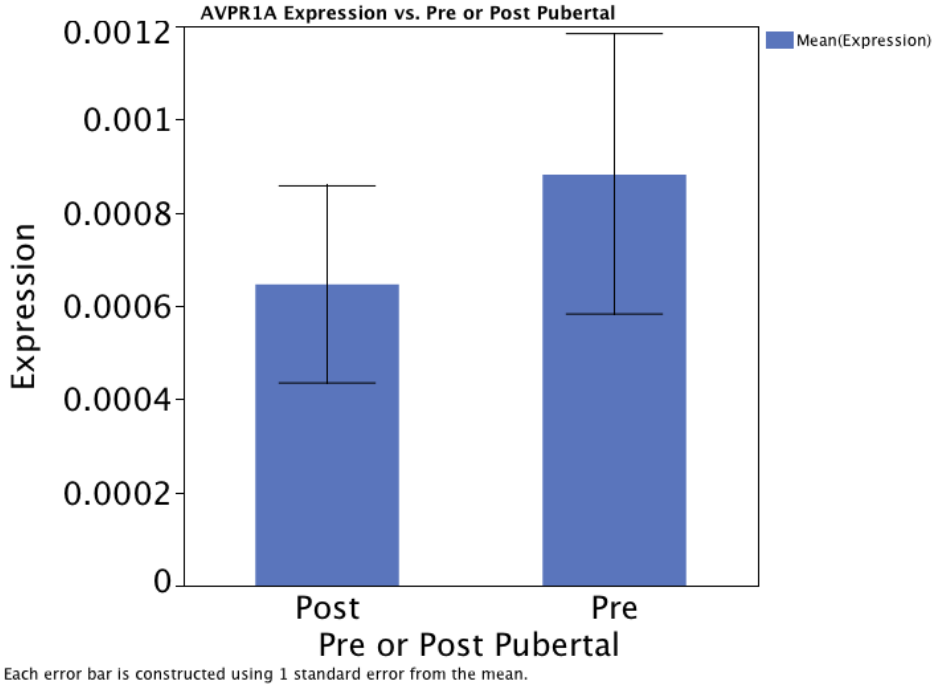


Figure 13 - *AVPR1A* gene expression: pre-pubertal gilt *G31E* genotype AA (n = 3) vs. AG (n = 5) in the anterior pituitary

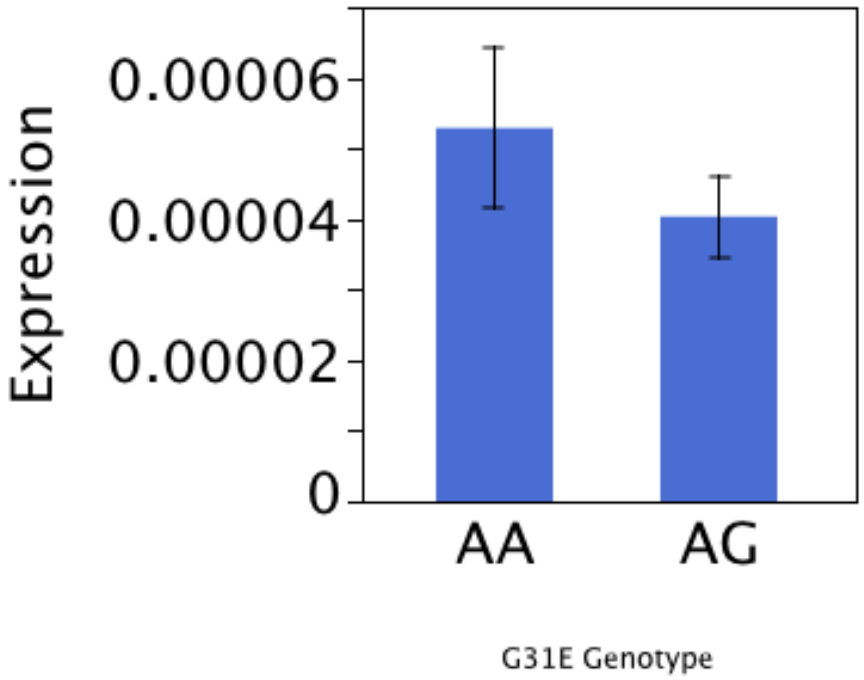
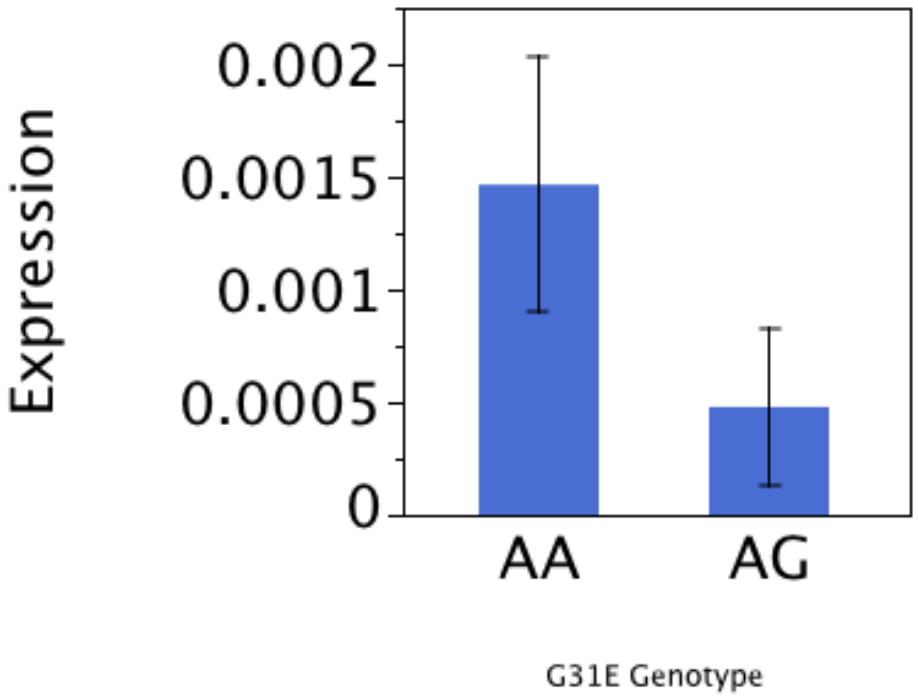


Figure 14 - *AVPR1A* gene expression: pre-pubertal gilt *G31E* genotype AA (n = 3) vs. AG (n = 4) in the posterior pituitary



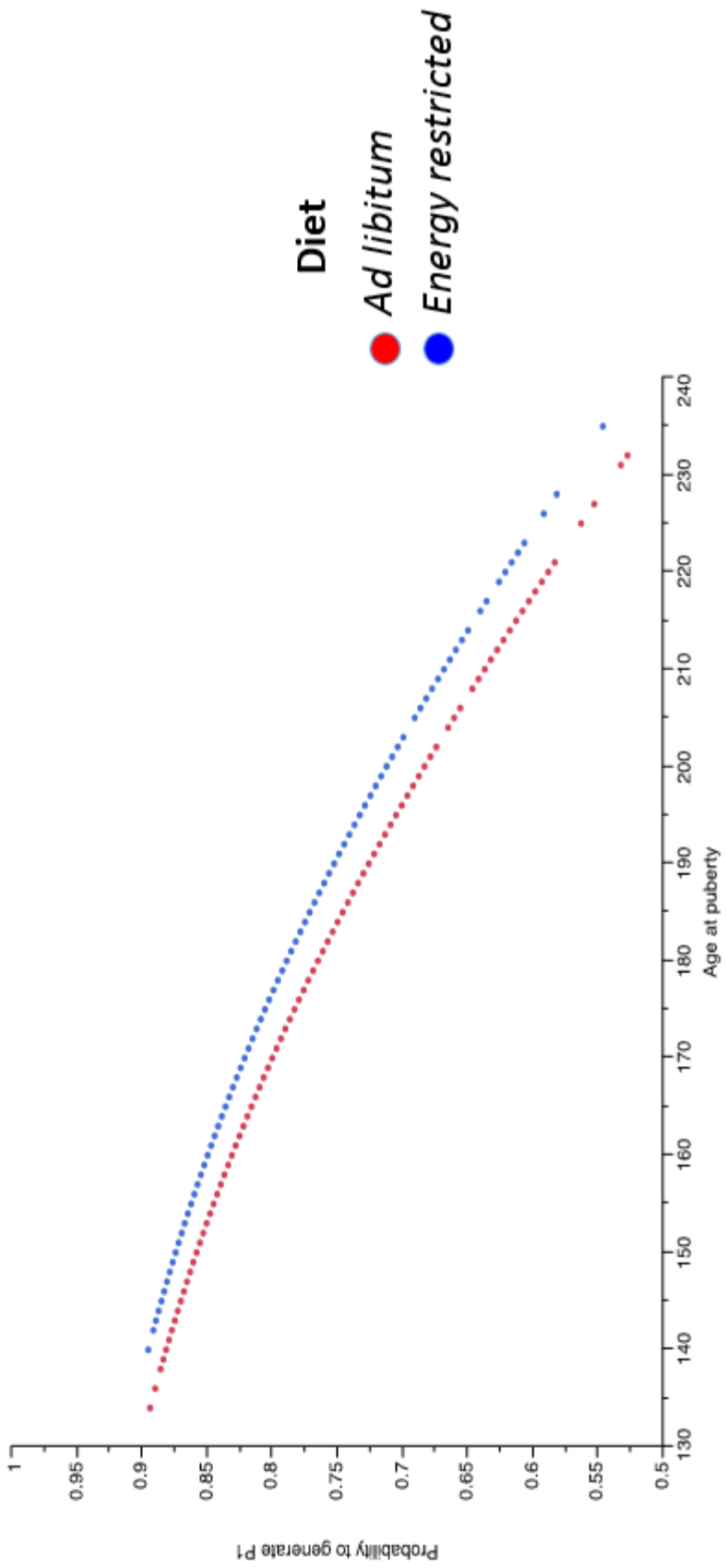


Figure 15 - Age at puberty and diet influence fertility

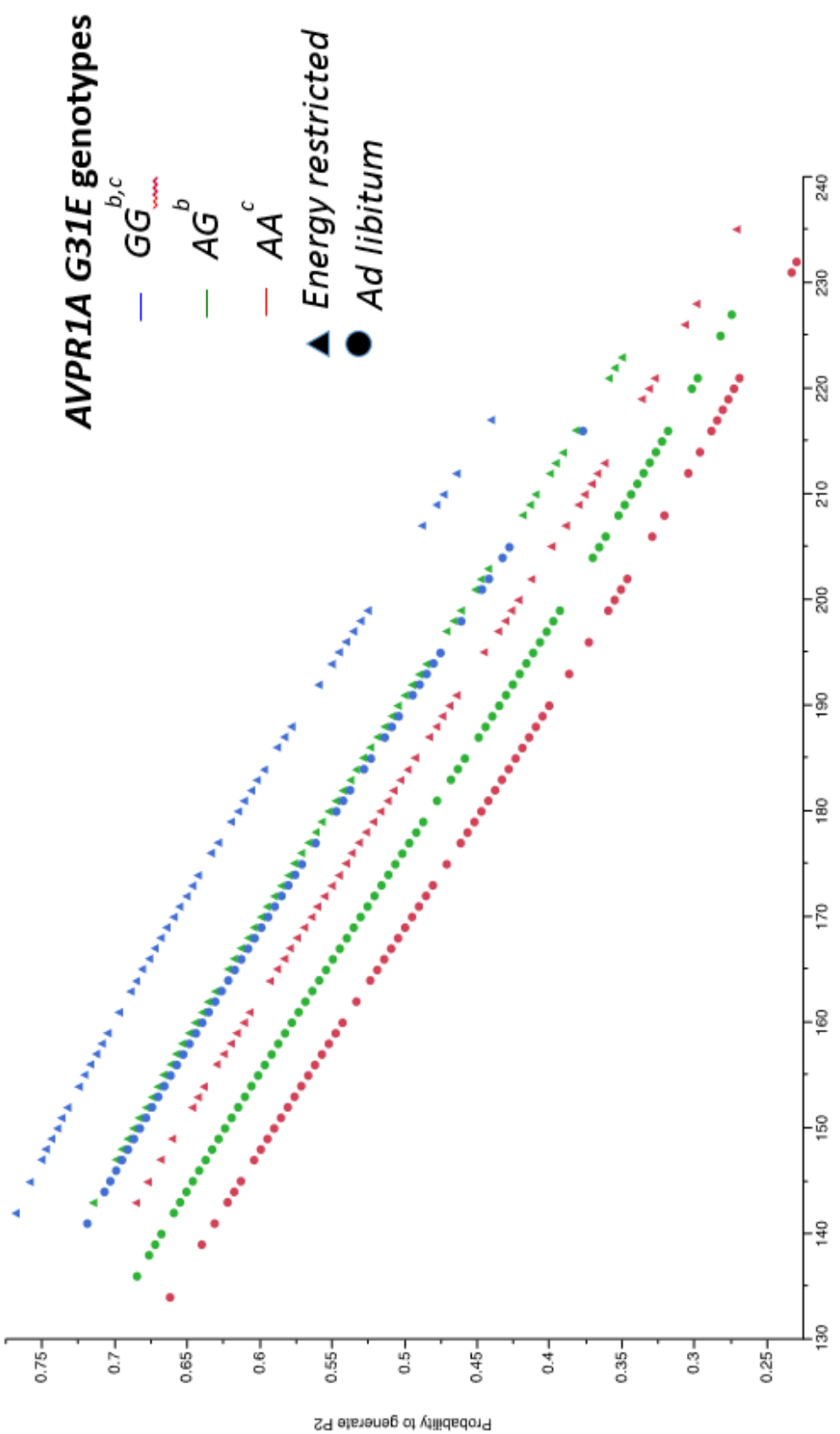
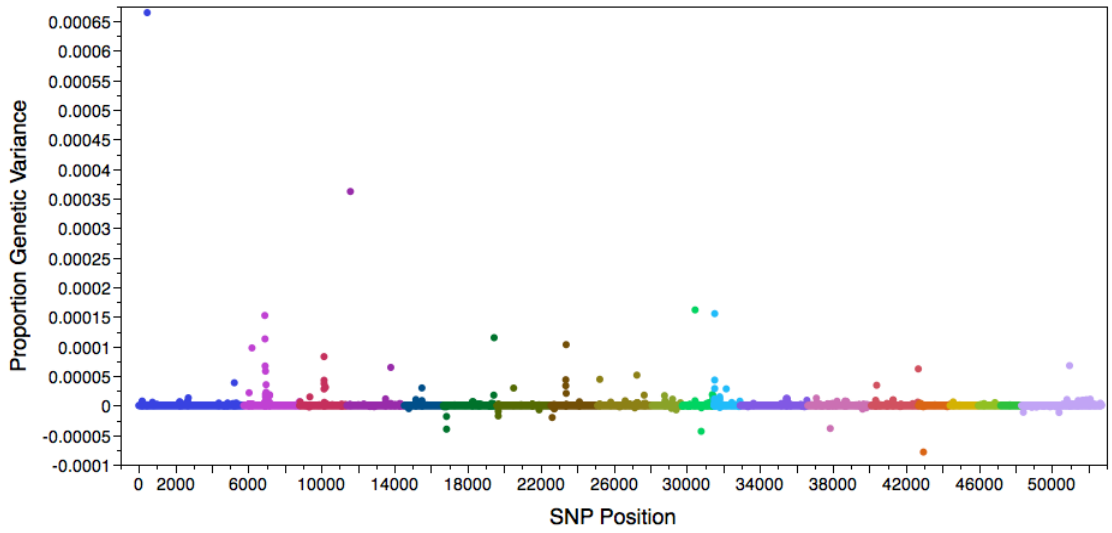


Figure 16 – *AVPR1A* genotypes interact with diet to influence the success rate of sows generating parity two

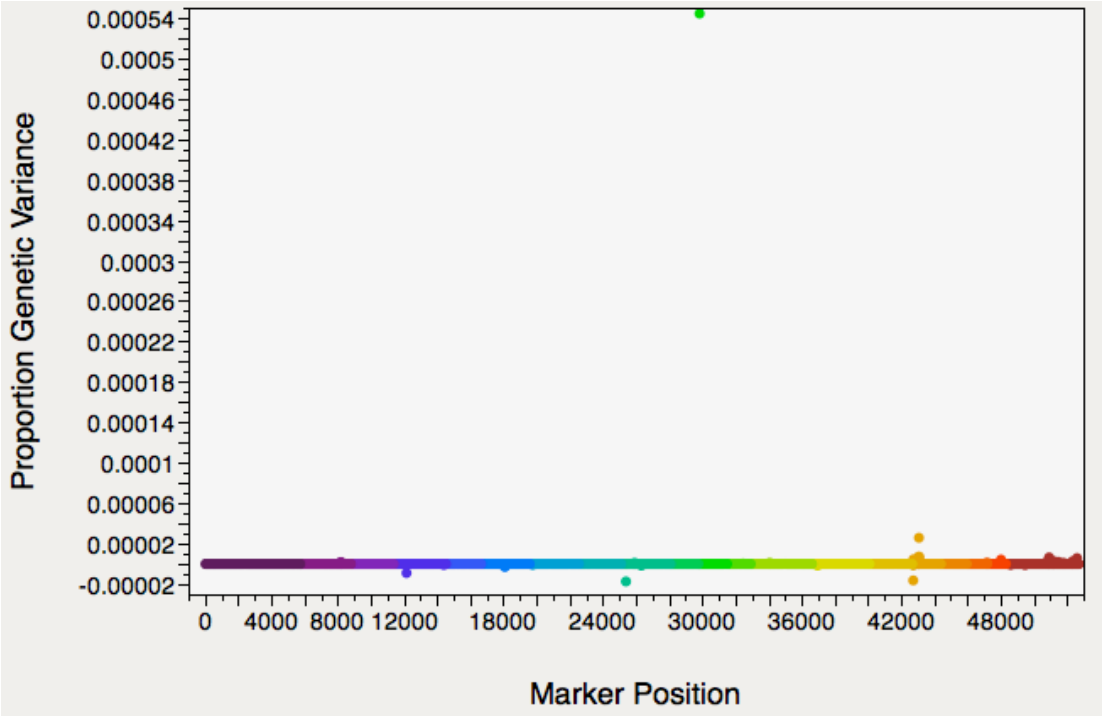
Genotypes with the same superscript differ: a, $P < 0.1$; b, $P < 0.05$; c, $P < 0.01$

Figure 17 – GWAS between 2 x 52,736 SNPs and AP



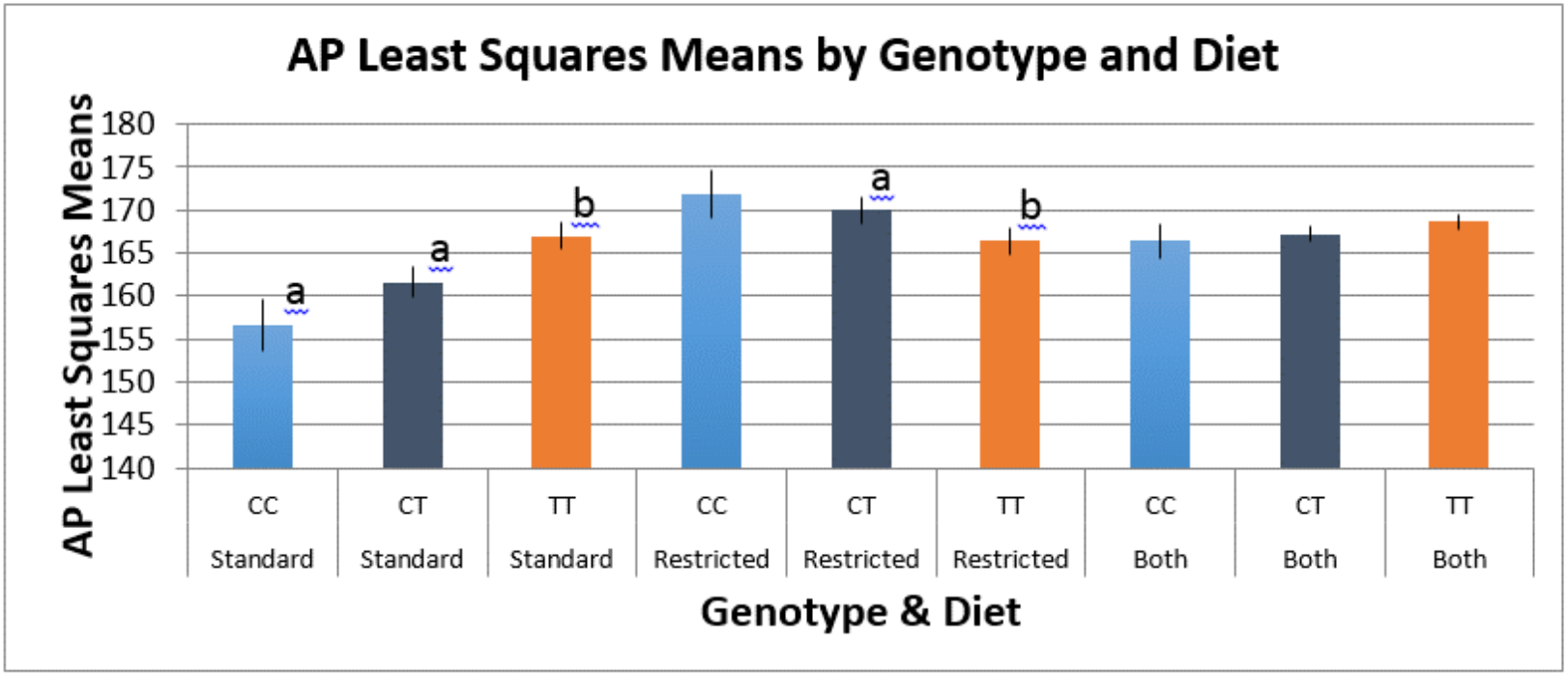
Each dot represents the proportion of genetic variance explained by SNPs. The X-axis represents the location of the SNPs in the swine genome. The Y-axis represents the contribution of that marker to the genetic variance. Positive values indicate associations due to genotype. Negative values indicate associations due to genotype by diet interaction. Alternate colors represent autosomes, from SSC1 to 18, followed by SSCX and SNPs with unknown location.

Figure 18 – GWAS between 2 x 52,736 SNPs and LTNP



Each dot represents the proportion of genetic variance explained by SNPs. The X-axis represents the location of the SNPs in the swine genome. The Y-axis represents the contribution of that marker to the genetic variance. Positive values indicate associations due to genotype. Negative values indicate associations due to genotype by diet interaction. Alternate colors represent autosomes, from SSC1 to 18, followed by SSCX and SNPs with unknown location.

Figure 19 – AP least squares means by genotype and diet of *MARC0053591*, a SNP identified by GWAS as interacting with diet to influence AP



Additive P < 0.0001
Dominance P < 0.05

Additive P < 0.01
Dominance P = 0.06

Additive P = 0.16
Dominance P = 0.36

LSM with different superscripts differ (P < 0.05)